

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ
ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ
ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Кваліфікаційна наукова
праця на правах рукопису

СУЛІМА СВІТЛАНА ВАЛЕРІЇВНА

УДК 621.391

ДИСЕРТАЦІЯ
«МЕТОДИ РЕКОНФІГУРАЦІЇ ОБЧИСЛЮВАЛЬНИХ РЕСУРСІВ
БАЗОВОЇ МЕРЕЖІ НА ОСНОВІ ТЕХНОЛОГІЇ ВІРТУАЛІЗАЦІЇ»

05.12.02 – телекомунікаційні системи та мережі

Подається на здобуття наукового ступеня кандидата технічних наук

Дисертація містить результати власних досліджень. Використання ідей,
результатів і текстів інших авторів мають посилання на відповідне джерело

(підпис, ініціали та прізвище здобувача)

Науковий керівник Скулиш Марія Анатоліївна, кандидат технічних наук,
старший науковий співробітник.

Київ – 2019

АНОТАЦІЯ

Суліма С.В. Методи реконфігурації обчислювальних ресурсів базової мережі на основі технології віртуалізації. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.12.02 – телекомунікаційні системи та мережі. – Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» МОН України, Київ, 2019.

Аналіз поточної ситуації в області безпроводового зв'язку показує збільшення службового навантаження, що призводить до збільшення необхідності у додаткових обчислювальних ресурсах, разом з тим нерівномірність завантаження вузлів інфраструктури призводить до їх простою, таким чином виникає потреба у впровадженні технологій, які зменшують простої обладнання і гарантують якість обслуговування навантаження на заданому рівні протягом дня. Щоб впоратися зі значним зростанням трафіку службових даних мобільний оператор освоює технології віртуалізації мережі та хмарних обчислень, щоб побудувати масштабовані та гнучкі мобільні мережі та пропонувати їх як хмарний сервіс.

Надання послуг в телекомунікаційній галузі традиційно ґрунтується на тому, що мережеві оператори впроваджують статичне фізичне обладнання для кожної мережевої функції, яке розміщується в фіксованих точках і має фіксовані характеристики обслуговування. На відміну від існуючої статичної архітектури мережі LTE EPC пропонується система, в якій виділені апаратні мережеві функції обслуговують певний заданий рівень службового навантаження, в той час як заявки службових потоків, що цей рівень перевищують, направляються на обробку з використанням віртуалізованих мережевих функцій у орендованих хмарах дата центрів.

Дисертаційну роботу присвячено вирішенню важливої науково-технічної проблеми, пов'язаної з розробкою наукових основ організації керування конфігурацією обчислювальних ресурсів в мережах оператора з

гібридною структурою, що характеризується наявністю як виділених спеціалізованих апаратних ресурсів, так і гнучких віртуалізованих обчислювальних ресурсів. Впровадження запропонованих методів дозволить забезпечити якісне обслуговування телекомунікаційних сервісів з мінімізацією капітальних, експлуатаційних витрат та забезпечить необхідний рівень еластичності мережі мобільного зв'язку. Розроблено структурований підхід до процесу керування ресурсами мережевих функцій за допомогою послідовного забезпечення контролю на етапах: моніторингу, прогнозування, контролю достатності ресурсів.

Виконано огляд літературних джерел за темою дисертації, аналіз принципів функціонування мережі мобільного зв'язку, особливостей процесу обслуговування сервісів в ньому. Проведено систематизацію методів керування обчислювальними ресурсами обслуговування в умовах гібридної інфраструктури телекомунікаційного оператора.

Перший розділ присвячено дослідженню та удосконаленню підходів щодо побудови обчислювально-телекомунікаційного середовища мережі мобільного зв'язку. Розроблено метод визначення місця розміщення та обсягів зарезервованих обчислювальних ресурсів віртуальних мережевих функцій, що враховує стан обчислювальних та мережевих ресурсів, вимоги до якості обслуговування службових потоків у ЕРС, гібридність мережі та на відміну від існуючих за рахунок динамічного надання ресурсів дозволяє зменшити обсяг використаних ресурсів, здійснюючи розрахунок топології та обсягу ресурсів одночасно. Визначивши цільову функцію задачі «вбудовування» віртуальних мережевих функцій і врахувавши доступні ресурси на вузлах та їх вартість, інтенсивність надлишкового службового навантаження та допустимі затримки, будується метод визначення топології виділення ресурсів та їх необхідного зарезервованого обсягу. Метод базується на попередньому визначенню за моделлю бінарного цілочисельного лінійного програмування та спільному розташуванні індивідуальних віртуалізованих сервісів базової мережі на фізичній мережі. Згідно з рішенням, кожній

мережевій функції резервується певна кількість ресурсів на основі оцінки її найбільшої потреби в ресурсах; миттєві потреби різних мережевих функцій динамічно задовольняються шляхом активації необхідної конфігурації віртуальних машин під час виконання таким чином, щоб задовольнити гарантії передбачені для кожної мережевої функції. Експериментальні дослідження показали зменшення умовних витрат в середньому до 15% у порівнянні з існуючими аналогами.

Другий розділ роботи вирішує завдання побудови системи керування обчислювальними ресурсами віртуалізованих мережевих функцій для надання гнучкості системі обслуговування. Запропоновано метод визначення розміру інтервалу часу сталої конфігурації обчислювальних ресурсів, який враховує витрати на реконфігурацію системи ЕРС та прогнозовані характеристики обслуговування телекомунікаційних сервісів, а також змінюючи розмір інтервалу перерозподілу обчислювальних ресурсів залежно від динаміки навантаження дозволяє зменшити інтенсивність проведення реконфігурації системи обслуговування та за рахунок перерозподілу обчислювальних ресурсів гарантувати задану якість обслуговування, підвищуючи коефіцієнт використання ресурсів.

Рішення про те, коли надавати ресурси, залежить від динаміки навантажень. Телекомунікаційні навантаження зазнають довгострокових змін, таких як вплив години дня або сезонні ефекти, а також короткострокових коливань таких як неочікувані навантаження. У той час як довгострокові коливання можуть бути передбачені заздалегідь, спостерігаючи за змінами в минулому, короткострокові коливання менш передбачувані, а в деяких випадках, не передбачувані. Запропонований метод використовує два різних підходи для роботи в умовах змін, які спостерігаються в різних часових масштабах. Використовується прогностичне керування ресурсами для оцінки навантаження і відповідного керування, а також реактивне керування ресурсами для виправлення помилок у довгострокових прогнозах або для реагування на непередбачені перенавантаження. Пропонується

застосовувати механізм, суть якого полягає у динамічній зміні тривалості сталої конфігурації ресурсів віртуальної мережевої функції залежно від різниці між максимальним значенням навантаження на певному базовому інтервалі та мінімальним на основі зібраних за довготривалий період часу даних стану елемента мережі. Якщо це значення збільшується, то скорочується інтервал сталої конфігурації мережевих функцій. Для роботи методу також потрібно визначати опорне базове значення інтервалу з урахуванням необхідності мінімізації похибки представлення навантаження з одного боку та числа інтервалів з іншого. Пропонується здійснювати постійний моніторинг значень інтенсивності навантаження та використовувати прогностичний метод, якщо ці значення не перевищують порогові, в іншому випадку – оцінювати поточні тренди у навантаженні та здійснювати масштабування ресурсів на основі нового прогнозу. Прогнозування навантаження на наступний інтервал часу здійснюється шляхом урахування довгострокової статистики та коригування його за моделлю експоненційного згладжування, де помилки прогнозів більш нових минулих періодів мають більший ваговий коефіцієнт. Представлений підхід дозволяє досягати раціонального співвідношення затрат на керування та підсумкового значення якості обслуговування. Результати моделювання системи з динамічною адаптацією величини інтервала та системи без неї показали, що різниця між прогнозованим значенням та репрезентаційним може складати 9%. Якщо не застосовувати систему динамічного регулювання величини вікна керування, то відхилення складатиме 18%, тобто на 9% більше і ресурсів буде витрачатись, відповідно, більше. Для оцінки запропонованого підходу середня кількість вільних ресурсів в день визначалась як різниця між фіксовано виділеними ресурсами, тобто коли протягом доби завжди виділено 100% ресурсів, та динамічно виділеними ресурсами з використанням NFV. Об'єм ресурсів, що виділяються динамічно, в середньому на 42% менше, ніж у разі використання традиційного підходу розподілу.

В четвертому розділі вирішено задачу щодо визначення місця міграції мережеских функцій з постраждалих внаслідок відмови або перевантаження вузлів, з одночасною оптимізацією вартості відновлення вузла після відмови, а також підтримкою високого рівня фізичної працездатності мережі. Потенційними джерелами відмов можуть бути: відмови фізичних вузлів, відмови серверів, які мають вищу інтенсивність відмов ніж телекомунікаційне апаратне забезпечення, або провайдер інфраструктури буде здійснювати задачі техобслуговування і це потребуватиме міграції вузлів. Розвинено метод локальної реконфігурації обчислювальних ресурсів віртуальної мережі у випадку відмови або перевантаження ЕРС, який враховує динамічну зміну топології мережі обслуговування та на відміну від існуючих дозволяє гнучко визначати конфігурацію обчислювальних ресурсів обслуговування, розміщати мережескі функції з мінімальними витратами та децентралізовано контролювати показники якості.

Пропонований підхід виконується розподіленим способом. У дослідженні вирішується проблема розташування децентралізованих вузлів керування з урахуванням параметрів затримки та стійкості, пропонується цільова функція комбінаторної оптимізаційної задачі розташування вузлів керування. Вибір вузлів-менеджерів здійснюється відповідно до критеріїв (1) мінімальної затримки до менеджера; (2) мінімального дисбалансу навантаження на менеджерів; (3) мінімальної затримки між менеджерами.

Метод локальної реконфігурації вузлів після перевантаження або відмови використовує евристичний підхід для визначення оптимального місця міграції віртуалізованих мережеских функцій. Коли вузол відмовляє, менеджер направляє запит на відновлення до всіх вузлів фізичної мережі, на яких розміщуються віртуальні вузли, суміжні з ураженими віртуальними вузлами; кожен з цих вузлів будує дерево найкоротших шляхів (SPT) до всіх вузлів фізичної мережі на відстані не більше порогової; менеджер обирає вузол з оптимальною відстанню до всіх вузлів фізичної мережі, де розташовані вузли віртуальної мережі прилеглі до несправного вузла.

Обираємо вузол з мінімальною вартістю шляху до всіх кореневих вузлів у деревах SPT та мінімальною вартістю обчислень. Якщо знайти новий вузол-кандидат для вузла віртуальної мережі, розміщеного на вузлі з відмовою, не вийде, то в такому випадку виконується процедура реконфігурації для міграції одного або декількох віртуальних вузлів для переміщення розміщених вузлів віртуальної мережі. Розглядаємо цю задачу як задачу оптимізації, яка спрямована на мінімізацію загальної вартості міграції при обмеженнях допустимої затримки і обчислювальних ресурсів. За результатами моделювання запропонований метод показав до 27% менші умовні витрати у порівнянні зі стратегією направленою на мінімізацію затримки, при цьому затримка знаходилась у допустимих межах але була на 20% більшою.

Таким чином, перед початком роботи необхідно мати статистичні дані інтенсивності надходження заявок на мережеву функцію та ймовірнісні характеристики обслуговування заявки. Згідно з методом розміщення визначається прив'язка кожної мережевої функції традиційної мережі до дата центру та кількості ресурсів, що потрібно зарезервувати для відповідної віртуалізованої мережевої функції. Далі необхідно розбити цикл роботи мережевої функції на інтервали протягом яких її конфігурація буде незмінною і відповідно буде активізовано визначений обсяг ресурсів згідно з методом визначення розміру інтервалу часу сталої конфігурації ресурсів з урахуванням очікуваного навантаження. Під час функціонування мережі мобільного зв'язку може виникнути ситуація, коли фізичний вузол не в змозі надалі обслуговувати вхідне навантаження через недостатність ресурсів для цього або через вихід його з ладу, в такому випадку запускається розподілений метод локальної реконфігурації ресурсів, який перерозподіляє віртуальні вузли у мережі, що постраждали від відмови або перевантаження на фізичному вузлі.

Проведені дисертаційні дослідження і отримані результати мають практичне значення при створенні блоку прийняття рішень щодо виділення

обчислювальних ресурсів системи ЕРС оператора мобільного зв'язку із повним або частковим застосуванням принципу віртуалізації мережевих функцій. Результати досліджень носять інтердисциплінарний характер, впроваджені у навчальний процес Інституту телекомунікаційних систем КПІ ім. Ігоря Сікорського. Проведено апробацію та виконано практичне тестування засобів керування перерозподілом ресурсів в лабораторіях телекомунікаційних компаній Huawei та Lifecell, результати моделювання запропонованих методів показали економію ресурсів обслуговування до 47%.

Ключові слова: мобільна мережа, базова мережа, віртуалізація, NFV, система керування навантаженням, конфігурація ресурсів, реконфігурація мережі.

ABSTRACT

Sulima S.V. Reconfiguration methods of the computing resources for the core network based on virtualization technology. – Qualification scientific paper, manuscript.

A thesis submitted in fulfilment of the Candidate of Sciences degree in technical sciences on specialty 05.12.02 – telecommunication systems and networks. – National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute» of Ministry of Education and Science of Ukraine, Kyiv, 2019.

The analysis of the current situation in the field of wireless communication shows an increase in the signaling traffic, that leads to an increase in the need for additional computing resources, however, the irregularity in the load of infrastructure nodes leads to their underload, thus there is a need for the introduction of technologies that reduce equipment underload and guarantee Quality of Service at a given level throughout the day. To cope with a significant increase in signaling traffic, mobile operator is developing network virtualization and cloud computing technologies to build scalable and flexible mobile networks and offer them as cloud service.

Provision of services in the telecommunication industry is traditionally based on the fact that network operators implement static physical equipment for each network function that is located at fixed points and has fixed service characteristics. Unlike the existing static architecture of the LTE EPC network, a system in which the dedicated physical network functions serve a given level of signaling traffic, while service flows requests that exceed this level are sent for processing using virtualized network functions in leased cloud data centers is offered.

The thesis is devoted to the solution of an important scientific and technical problem related to the development of the scientific basis for managing the configuration of computing resources in networks of an operator with a heterogeneous structure, characterized by the presence of both dedicated hardware resources and flexible virtualized computing resources. The introduction of the proposed methods will allow providing of high-quality telecommunication services with minimization of capital expenditures, operational expenditures and providing the necessary level of elasticity of the mobile communication network. A structured approach to the process of network functions resource management is developed based on successive control at the stages of monitoring, forecasting, controlling the sufficiency of resources.

The review of literary sources on the topic of the thesis, analysis of the functioning principles of the mobile communication network, peculiarities of the services processing in it are done. The systematization of methods for computing resources management in the conditions of telecommunication operator heterogeneous infrastructure is carried out.

The first section is devoted to the research and development of approaches to computing-telecommunication environment of the mobile communication network construction. The method of determining the location and capacity of reserved computing resources of virtual network functions, taking into account the state of computing and network resources, requirements to the Quality of Service for signaling flows in the EPC, heterogeneity of the network and, unlike existing

ones due to the dynamic provision of resources, can reduce the amount of resources used, implementing calculation of topology and capacity of resources at the same time is developed. Having determined the objective function of the task of virtual network functions «embedding» and by taking into account available resources at the nodes and their cost, the intensity of excessive signaling traffic and admissible delays, a method for determining the topology of resource allocation and their required reserved capacity is developed. The method is based on a preliminary definition based on the binary integer linear programming model and the joint embedding of the individual virtualized core network services on the physical network. According to the solution, for each network function a certain amount of resources is reserved based on the assessment of its biggest resource requirements; the instantaneous needs of different network functions are dynamically satisfied by activating the required virtual machine configuration during execution in such a way as to satisfy the guarantees provided for each network function. Experimental studies have shown a reduction in conventional costs by an average of up to 15% compared to existing prototypes.

The second section of the work solves the task of building a virtualized network functions computing resources management system to provide flexibility to the serving system. The method of determining the size of the computing resources constant configuration time interval, which takes into account the costs of reconfiguring the EPC system and the predicted characteristics of the telecommunication services processing, and also by changing the size of the computing resources reallocation interval according to the traffic dynamics allows reducing the intensity of the service system reconfiguration and by reallocation of computing resources allows guaranteeing the given Quality of Service, increasing the coefficient of resources usage is proposed.

The decision on when to provide resources depends on the dynamics of traffic. Telecommunication traffic experience long-term changes, such as daylight hours or seasonal effects, as well as short-term fluctuations such as unexpected crowds. While long-term fluctuations can be predicted in advance, by observing

changes in the past, short-term fluctuations are less predictable, and in some cases, unpredictable. The proposed method uses two different approaches to work in conditions of changes that are observed at different time scales. Prognostic resource management is used to evaluate the load and appropriate management, as well as reactive resource management is used to correct long-term errors or to respond to unforeseen overload. It is proposed to apply a mechanism that dynamically changes the duration of the virtual network function resources constant configuration depending on the difference between the maximum traffic value on a certain base interval and the minimum value based on the network element's state data collected for a long time period. If this value increases, then the interval of constant configuration of network functions is reduced. In order to work the method also requires determining the reference base value of the interval, taking into account the need to minimize the traffic representation error on one side and the number of intervals on the other. It is proposed to continuously monitor the traffic intensity values and use the prognostic method if these values do not exceed the threshold; otherwise, to assess the current trends in traffic and scale resources based on the new forecast. Traffic forecasting for the next time interval is carried out by taking into account long-term statistics and adjusting it according to the model of exponential smoothing, where the forecasting errors of newer past periods have more weight. The presented approach allows achieving a rational ratio of management costs and the resulting Quality of Service. Modeling results of the system with dynamic adaptation of the interval value and the system without it showed that the difference between the predicted value and the representative can be 9%. If you do not apply the dynamic adjustment system of the control window value, then the deviation will be 18%, that is, 9% more and the resources will be spent, respectively, more. To assess the proposed approach, the average amount of free resources per day was defined as the difference between fixed allocations, i.e., when 100% of resources were always allocated during the day, and dynamically allocated resources using NFV. The volume of resources allocated dynamically, on average, is 42% less than in the case of using the traditional allocation approach.

The fourth section solves the problem of determining the network functions migration location from the nodes affected by failure or overload, while optimizing the cost of recovery of the node after failure, as well as maintaining a high level of physical performance of the network. Potential failures can be: physical nodes failures, failures of servers that have higher failure rates than telecommunication hardware, or the infrastructure provider will perform maintenance tasks and this will require the migration of nodes. Method of virtual network computing resources local reconfiguration in the case of failure or overload of the EPC, that takes into account the dynamic change in the topology of the serving network, and unlike the existing ones, allows flexible determination of processing resources configuration, placing network functions with minimal costs, and decentralized monitoring of quality indicators is developed.

The proposed approach is implemented in a distributed manner. The research solves the problem of the location of decentralized control nodes with allowance for latency and stability parameters, and the target function of the combinatorial optimization problem of the control nodes location is proposed. The choice of manager nodes is carried out according to the criteria (1) of minimal delay to the manager; (2) a minimum imbalance of the load on managers; (3) minimum delay between managers.

The method of nodes local reconfiguration after an overload or failure uses a heuristic approach to determine the optimal location for migration of virtualized network functions. When the node fails, the manager sends the recovery request to all nodes of the physical network, which hosts the virtual nodes adjacent to the affected virtual nodes; each of these nodes builds the shortest path tree (SPT) to all nodes of the physical network at a distance not exceeding the threshold; the manager selects the node with the optimal distance to all nodes of the physical network, where the virtual network nodes adjacent to the faulty node are located. Node with the minimum cost of the path to all root nodes in the SPT and the minimum cost of the processing is selected. If a new host node for a virtual node located on a failure node can not be found, then a reconfiguration procedure is

performed to migrate one or more virtual nodes to move the placed virtual network nodes. We consider this task as an optimization problem, which is aimed at minimizing the total cost of migration with the limits of permissible delay and computational resources. According to the simulation results, the proposed method showed up to 27% less conventional costs compared to a strategy aimed at minimizing delay, while the delay was within the permissible limits but was 20% higher.

Thus, before the start of work it is necessary to have statistics on the input arrival rate for the network function and the probabilistic characteristics of request processing. According to the allocation method, the binding of each network function of the traditional network to the data center and the number of resources to be reserved for the corresponding virtualized network function is determined. Next it is necessary to break the network function operation cycle at intervals during which its configuration will remain unchanged and accordingly a certain amount of resources will be activated by the method of determining the resources constant configuration time interval size, taking into account the expected traffic load. During the operation of the mobile network, a physical node may not be able to maintain the input traffic load due to lack of resources for this or its failure, in which case a distributed method of local reconfiguration of the resources of that reallocates virtual nodes in the network, that suffered from a failure or overload on a physical node.

Thesis researches and obtained results have practical significance when creating a decision block for allocation of EPC system computing resources of the mobile communication operator with full or partial application of the principle of network functions virtualization. The research results are interdisciplinary, implemented in the educational process of the Igor Sikorsky Kyiv Polytechnic Institute Institute of Telecommunication Systems. Approbation and practical testing of resource allocation management tools have been carried out in Huawei and Lifecell telecommunication laboratories, the results of the simulation of the proposed methods have shown savings of up to 47% of service resources.

Keywords: mobile network, core network, virtualization, NFV, traffic management system, resource configuration, network reconfiguration.

Список публікацій здобувача:

Монографія

1. Математичні методи аналізу та керування телекомунікаційними мережами. Розділ 5 // Л.С. Глоба, О.М. Дяденко, А.Ю. Пилипенко, М.А. Скулиш. –К.: Інститут обдарованої дитини НАПН України. — ISBN 978-966-2633-84-9. — 2018 . — С. 206-225.

Особистий внесок: система керування відображенням та плануванням ресурсів телекомунікаційної мережі з використанням технології віртуалізації.

Статті у наукових фахових виданнях

2. Скулиш М. А. Організація управління в багатоетапних системах масового обслуговування / М. А. Скулиш, С. В. Суліма // Вісник Національного університету «Львівська політехніка». Радіoeлектроніка та телекомунікації. — 2015. — № 818. — С. 80–85.

Особистий внесок: аналітична модель для визначення кількості ресурсів, яку необхідно виділити для кожного етапу багатоетапної системи.

3. Skulysh M. Model for Efficient Allocation of Network Functions in Hybrid Environment / M. Skulysh, L. Globa, S. Sulima // Information and Telecommunication Sciences. — 2016. — № 1. — С. 39–45.

Особистий внесок: модель виділення ресурсів системи даних центрів для визначення оптимальної кількості ресурсів виділених мережевому функціональному вузлу.

4. Суліма С. В. Метод відновлення мережі у віртуалізованому середовищі / С. В. Суліма // Радіoeлектроніка та інформатика. — 2017. — № 4. — С. 4–8.

5. Skulysh M. Hybrid resource management system for telecommunication network / M. Skulysh, S. Sulima // Advanced Information Systems. — 2018. — Vol. 2, № 1. — P. 47–51.

Особистий внесок: спосіб керування ресурсами телекомунікаційної мережі шляхом вибору оптимального інтервалу для перерозподілу системних ресурсів.

Статті у фахових виданнях України, які включені до міжнародних наукометричних баз

6. Суліма С. В. Система управління ресурсами в центрах обробки даних оператора мережі мобільного зв'язку / С. В. Суліма, М. А. Скулиш // Вісник НТУУ «КПІ». Серія Радіотехніка, Радіоапаратобудування. — 2017. — № 68. — С. 27-32. (Наукове фахове видання України з технічних наук. Входить до міжнародних наукометричних баз: DOAJ, Web of Science, Index Copernicus, eLibrary.ru / РИНЦ та ін.)

Особистий внесок: метод динамічної конфігурації ресурсів.

7. Суліма С. В. Гібридна система управління ресурсами для віртуалізованих мережевих функцій / С. В. Суліма, М. А. Скулиш // Радіоелектроніка, інформатика, управління.— 2017. — № 1(40). — С. 16–23. (Наукове фахове видання України з технічних наук. Входить до міжнародних наукометричних баз: Web of Science, DOI, DOAJ, British Library, eLibrary.ru / РИНЦ, Index Copernicus та ін.)

Особистий внесок: метод побудови системи управління ресурсами для віртуальних мережевих функцій в центрах обробки даних оператора мережі мобільного зв'язку.

Статті у інших виданнях

8. Скулиш М. А. Метод управління ресурсами в датацентрах оператора мережі мобільного зв'язку / М. А. Скулиш, С. В. Суліма // Радіоелектроніка та інформатика. — 2015. — № 3. — С. 8–13.

Особистий внесок: модель для визначення оптимальної кількості датацентрів, топології їх розміщення та кількості ресурсів, які виділяються

функціональним блокам системи, що розгортаються в дата центрах та передбачають використання декількох типів ресурсів.

Тези наукових доповідей

9. Skulysh M. Management of Multiple Stage Queuing Systems / M. Skulysh, S. Sulima // CADSM 2015 : 13-th International conference, 24–27 February 2015 : IEEE Digital Library. — Lviv–Polyana, 2015. — P. 431–433.

Особистий внесок: модель динамічного виділення ресурсів у багатоступінних системах.

10. Скулиш М. А. Організація управління в багатоступінних системах масового обслуговування / М. А. Скулиш, С. В. Суліма // Проблеми телекомунікацій ПТ-2015 : 9-а міжнародна науково-технічна конференція, 21–24 квітня 2015 : матеріали конференції. — Київ, 2015. — С. 319–321.

Особистий внесок: метод динамічного керування ресурсами багатоступінних систем.

11. Скулиш М. А. Управління ресурсами в віртуалізованих мережах оператора мобільного зв'язку / М. А. Скулиш, С. В. Суліма // Інфокомунікації – сучасність та майбутнє : 5-а міжнародна науково-практична конференція, 29-30 жовтня 2015 : матеріали конференції. — Одеса, 2015. — С. 84–88.

Особистий внесок: метод початкової організації гетерогенних ресурсів мережі.

12. Globa L. Method for resource allocation of virtualized network functions in hybrid environment / L. Globa, M. Skulysh, S. Sulima // 2016 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), 6–9 June 2016 : IEEE Digital Library. — Varna, Bulgaria, 2016. — P. 1–5.

Особистий внесок: метод адаптивної конфігурації гетерогенних ресурсів мережі.

13. Суліма С. В. Алгоритм відображення та планування віртуалізованих функцій в мережі мобільного зв'язку / С. В. Суліма, М. А. Скулиш // Проблеми телекомунікацій ПТ-2016 : 10-а міжнародна науково-технічна

конференція, 19–22 квітня 2016 : матеріали конференції. — Київ, 2016. — С. 372–374.

Особистий внесок: метод відображення та планування віртуалізованих мережесевих функцій.

14. Skulysh M. Service deployment aspects in the systems with network function virtualization / M. Skulysh, S. Sulima // Radio Electronics & Info Communications : First International Conference, 11–16 September 2016 : IEEE Digital Library. — Kyiv, 2016. — P. 1–7.

Особистий внесок: підхід до моделювання та управління розподілом ресурсів віртуалізованих мережесевих функцій у мережі телекомунікаційного оператора.

15. Sunduchkov K. Sampling for direct search method of all system implementations developed according to the requirements of the technical design specification / K. Sunduchkov, M. Skulysh, S. Sulima, B. Savchuk // 2017 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), 10–13 October 2017 : IEEE Digital Library. — Kharkov, 2017. — P. 83–86.

Особистий внесок: формалізація модифікованого методу оптимізації побудови системи зв'язку.

16. Суліма С. В. Метод відновлення вузла у мережах NFV / С. В. Суліма // Проблеми телекомунікацій ПТ-2017 : 11-а міжнародна науково-технічна конференція, 18–21 квітня 2017 : матеріали конференції. — Київ, 2017. — С. 270–272.

17. Sulima S.V. Model for allocation of network slices in heterogeneous environment / S. V. Sulima // Modern challenges in telecommunications : Twelfth International Scientific Conference, 16–20 April 2018 : conference proceedings. — Kyiv, 2018. — С. 42–44.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	20
ВСТУП.....	22
РОЗДІЛ 1 Конфігурація ресурсів мережевих функцій в гібридному середовищі	29
1.1 Проблематика віртуалізації мережевих ресурсів	29
1.2 Огляд архітектури LTE/EPC	42
1.3 Якість обслуговування та EPS канали	48
1.4 Віртуалізація Мережевих Функцій (NFV)	49
1.4.1 Високорівнева платформа NFV	49
1.4.2 Функції та сервіси віртуальної мережі	54
1.4.3 NFV Управління та Оркестровка (NFV Management and Orchestration – NFV MANO)	56
1.5 Хмарні обчислення	57
1.5.1 Загальна характеристика.....	57
1.5.2 Взаємозв'язок між хмарними обчисленнями та NFV	59
1.5.3 NFV, SDN та хмарні обчислення	60
1.6 Опис мережевого сервісу.....	62
1.7 Особливості навантаження та роботи дата центра	67
1.8 Особливості дослідження.....	70
Висновки	78
РОЗДІЛ 2 Резервування ресурсів для віртуалізованих мережевих функцій в гібридному середовищі.....	80
2.1 Процедури LTE та потік заявок.....	80
2.2 Визначення кількості сервісів віртуальних мереж.....	87
2.2.1 Постановка задачі вибору вузлів агрегації трафіку	88
2.2.2 Формулювання оптимізаційної задачі	90
2.3 Відображення віртуальних вузлів на фізичні вузли.....	94
2.4 Еластичне керування EPC за допомогою поділу стану.....	99

2.4.1	Процедура встановлення інтерфейсу синхронізації	102
2.4.2	Процедура повторної синхронізації	103
2.5	Оркестрування віртуалізованої мережі EPC.....	104
2.6	Оцінка методу визначення місця розміщення та необхідної ємності віртуальних ресурсів у разі виникнення перевантаження фізичної мережі	107
	Висновки	109
РОЗДІЛ 3 Система керування ресурсами віртуалізованих мережевих функцій		110
3.1	Постановка задачі	110
3.2	Система динамічного керування ресурсами.....	112
3.3	Моніторинг та керування.....	114
3.4	Визначення інтервалу часу сталої конфігурації ресурсів	116
3.5	Розподілення ресурсів мережевих функцій	121
3.6	Метод прогнозування навантаження	121
3.7	Оцінка часу обслуговування.....	124
3.8	Аналіз методу керування ресурсами в контексті зміни інтенсивності надходження запитів	124
3.9	Експериментальне дослідження використання ресурсів у віртуалізованій базовій мобільній мережі.....	127
3.10	Автоматизація процесу масштабування мережевих функцій....	128
	Висновки	132
РОЗДІЛ 4 Реконфігурація мережі після перевантаження або збою		134
4.1	Модель мережі та постановка задачі	134
4.2	Оптимальне розміщення вузлів керування у мережах, заснованих на NFV	135
4.2.1	Затримка вузол-менеджер	137
4.2.2	Збалансований розподіл навантаження менеджерів	138
4.2.3	Затримка менеджер-менеджер.....	138
4.2.4	Цільова функція оптимізації.....	139

4.2.5	Конфігурація вузлів менеджерів	139
4.3	Схема відновлення мережі після відмови.....	140
4.4	Схема відновлення у разі перевантаження вузла	141
4.5	Реконфігурація перевантажених мереж.....	143
4.6	Оцінка методу локальної реконфігурації мережі	146
4.7	Модифікована система керування ресурсами.....	146
	Висновки	149
	ЗАГАЛЬНІ ВИСНОВКИ	150
	СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ.....	152
	ДОДАТОК А. Список публікацій здобувача за темою дисертації та відомості про апробацію результатів дисертації	167
	ДОДАТОК Б. Акти впровадження результатів дисертаційної роботи...	171

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

APN	Access Point Name
ARP	Allocation and Retention Priority
BSS	Business Support System
CN	Central Node
DRPA	Dynamic Resource Provisioning Algorithm
EM	Element Manager
EN	Edge Node
EPC	Evolved Packet Core
EPCaaS	EPC as a Service
EPS	Evolved Packet System
GBR	Guaranteed Bit Rate
HSM	Home Synchronization Manager
HSS	Home Subscriber System
IL	Instantiation Level
IMS	IP Multimedia Subsystem
IoT	Internet of Things
LEPC	LTE EPC
LTE	Long Term Evolution
MANO	Management and Orchestration
MCE	Management Component Entity
MME	Mobility Management Entity
NAS	Non Access Stratum
NFV	Network Functions Virtualization
NFVI	NFV Infrastructure
NFVO	NFV Orchestrator
NS	Network Service
NSD	Network Service Descriptor
OSS	Operations Support System

PCRF	Policy Control and Charging Rules Function
PDN	Packet Data Network
PGW	Packet Data Network Gateway
PSM	Proxy Synchronization Manager
QCI	QoS Class Identifier
QoS	Quality of Service
SDN	Software Defined Networking
SGW	Serving Gateway
SLA	Service Level Agreement
SPT	Shortest Path Tree
SSM	Serving Synchronization Manager
TAP	Traffic Aggregation Point
TSP	Telecommunication Service Provider
UE	User Equipment
VCD	Virtual Compute Descriptor
VDU	Virtualization Deployment Units
vEPC	Virtual EPC
VIM	Virtualized Infrastructure Manager
VL	Virtual Link
VLD	Virtual Link Descriptor
VNE	Virtual Network Embedding
VNF	Virtual Network Function
VNFC	Virtualised Network Function Component
VNFD	VNF Descriptor
VNFFG	VNF Forwarding Graph
VNFFGD	VNFFG Descriptor
VNFM	VNF Manager
VSD	Virtual Storage Descriptor

ВСТУП

Актуальність теми. У мережі мобільного стільникового зв'язку останнім часом спостерігається швидкий розвиток. Сьогоднішні мобільні мережі здатні запропонувати більш високі швидкості передачі даних, з інтеграцією більшої кількості послуг і гарантією високої якості сприйняття. Тим не менш, цей розвиток означає також, що обсяг даних, які передаються в мобільній мережі, також зростає, відповідно зростає обсяг службового трафіку. Крім того, конкуренція на ринку вимагає більш швидкого розгортання послуг і еластичність в зміні критеріїв обслуговування і здатність справлятися з більш високими вимогами обслуговування. Тому виникає необхідність у ефективному управлінні службовим трафіком з метою забезпечення потрібної якості обслуговування кінцевих користувачів та оптимального використання ресурсів мережі оператора зв'язку.

Поява концепції віртуалізації мережевих функцій відкриває нові можливості для світу телекомунікаційних систем, в той же час виникає необхідність у нових підходах, моделях та методах організації процесу обслуговування сервісів. Використання віртуальних серверів для вирішення задач базової мобільної мережі може значно спростити процес розподілу ресурсів на сервері обслуговування, забезпечить його масштабованість, підвищить відмовостійкість.

Вітчизняні та закордонні вчені пропонують нові архітектурні рішення для впровадження концепцій віртуалізації мережевих функцій з використанням пулу обчислювальних та мережевих ресурсів для побудови мереж всіх рівнів в тому числі базової мобільної мережі.

Невирішеним залишається ряд проблем. Необхідно врахувати гібридність обслуговуючого середовища, де присутні, по-перше, спеціалізовані апаратні засоби для надання послуг, які мають фіксовані характеристики, а отже обмежений ресурс, що ускладнює процес масштабування та реконфігурації мережі мобільного зв'язку; а, по-друге,

віртуалізовані ресурси з налаштовуваними характеристиками. Постає проблема інтеграції нових систем з існуючими апаратними рішеннями, тому запропоновані у роботі методи розподілу функціоналу обслуговування між віртуальними сутностями та існуючими апаратними засобами є актуальною задачею. На відміну від існуючої статичної архітектури мережі LTE EPC пропонується система, в якій виділені апаратні мережеві функції обслуговують певний заданий рівень службового навантаження, в той час як заявки службових потоків, що цей рівень перевищують, направляються на обробку з використанням віртуалізованих мережевих функцій у орендованих хмарах дата центрів. Після організації обчислювально-телекомунікаційного середовища виникає необхідність у адаптації системи протягом функціонування для реагування на зміни у навантаженні. При цьому слід мінімізувати додаткові витрати пов'язані з процесом керування, а отже вирішувати питання коли і як здійснювати реконфігурацію ресурсів обслуговування. В попередніх дослідженнях в контексті задачі керування ресурсами різні аспекти не розглядаються разом, а важливо розробити комплексний підхід, який дозволяє керувати ресурсами дата центрів гібридного середовища з урахуванням особливостей навантаження та мережних елементів.

Таким чином, науково-технічна задача розгортання та функціонування комунікаційних мереж на основі технології віртуалізації є актуальною.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційну роботу виконано згідно з планом наукових досліджень кафедри інформаційно-телекомунікаційних мереж КПІ ім. Ігоря Сікорського: науково-дослідної роботи «Технологія обробки сервісів з інтеграцією інформаційних ресурсів в системах підтримки операційної діяльності підприємств зв'язку» (номер державної реєстрації 0115U000217); ініціативної науково-дослідної теми «Технологія забезпечення якості обслуговування телекомунікаційних сервісів при частковому руйнуванні фізичного обладнання» (номер державної реєстрації 0117U001861).

Мета та задачі дослідження. Метою дослідження є збільшення коефіцієнту використання обчислювальних ресурсів під час функціонування базової мобільної мережі (ЕРС) для гарантування належного рівня якості надання гібридних сервісів за рахунок динамічного перерозподілу службових потоків в мережі та реконфігурації обчислювальних ресурсів обслуговування.

Об'єкт дослідження – процес керування обчислювальними ресурсами базової мобільної мережі LTE з гібридною структурою.

Предмет дослідження – методи та моделі обробки телекомунікаційних сервісів в мережах з гібридною інфраструктурою (структура яких складається з віртуалізованих обчислювальних ресурсів та спеціалізованих ресурсів оператора зв'язку).

Для досягнення мети дослідження було поставлено та вирішено такі *основні задачі*:

1. Проаналізувати принципи функціонування базової мережі мобільного зв'язку, особливості процесу обслуговування сервісів в ній, систематизувати методи керування обчислювальними ресурсами обслуговування в умовах гібридної інфраструктури телекомунікаційного оператора.

2. Розробити метод визначення місця розміщення та необхідної ємності віртуальних зарезервованих обчислювальних ресурсів у разі виникнення перевантаження фізичної мережі для забезпечення процесу обслуговування службових потоків у базовій мережі на належному рівні якості та підвищення коефіцієнту використання ресурсів системи, який не погіршуватиме показники обслуговування та мінімізуватиме обсяг необхідних обчислювальних ресурсів під час надання гібридних телекомунікаційних сервісів.

3. Розробити метод визначення розміру інтервалу часу сталої конфігурації ресурсів з урахуванням очікуваного навантаження в мережах з гібридною структурою, який за рахунок динамічної зміни розміру інтервалу керування ресурсами обслуговування телекомунікаційних сервісів дає змогу

зменшити кількість реконфігурацій системи обслуговування та за рахунок перерозподілу обчислювальних ресурсів гарантувати задану якість обслуговування.

4. Вдосконалити метод локальної реконфігурації обчислювальних ресурсів в мережі оператора мобільного зв'язку з гібридною структурою у разі виникнення нештатної ситуації за рахунок гнучкого вибору конфігурації розміщення мережеских функцій та децентралізованого контролю показників якості.

5. Провести оцінку запропонованих рішень щодо керування обчислювальними ресурсами в нормальних та нештатних умовах функціонування системи базової мобільної мережі.

Методи дослідження. Методи системного аналізу використовуються для синтезу системи керування ресурсами, визначення структурних взаємозв'язків її компонентів та поведінки, методи математичного аналізу, теорії масового обслуговування та теорії оптимізації використовуються для визначення місць розташування мережеских функцій та часу і кількості виділених їм ресурсів, математичне та імітаційне моделювання використовуються для дослідження роботи запропонованих методів реконфігурації ресурсів в гібридних мережах.

Наукова новизна одержаних результатів.

1. Вперше розроблено метод визначення місця розміщення та обсягів необхідних обчислювальних ресурсів віртуальних мережеских функцій, який на відміну від існуючих враховує доступні фізичні мережескі ресурси та віртуалізовані обчислювальні ресурси мережі, вимоги до якості обслуговування службових потоків у ЕРС, що дозволяє зменшити обсяг необхідних затрат ресурсів, здійснюючи розрахунок топології та обсягу ресурсів в одній моделі.

2. Вперше запропоновано метод визначення розміру інтервалу часу сталої конфігурації обчислювальних ресурсів, який на відміну від існуючих враховує витрати на реконфігурацію системи ЕРС та забезпечення заданих

показників обслуговування прогнозованого навантаження, а також подальшу оптимізацію інтервалу перерозподілу обчислювальних ресурсів залежно від динаміки навантаження, що дозволяє зменшити частоту проведення реконфігурації системи обслуговування та, за рахунок перерозподілу обчислювальних ресурсів, гарантувати задану якість обслуговування і підвищити коефіцієнт використання ресурсів.

3. Набув подальшого розвитку метод локальної реконфігурації обчислювальних ресурсів віртуальної мережі у випадку відмови або перевантаження підсистем ЕРС, який враховує динамічну зміну топології мережі обслуговування та, на відміну від існуючих, дозволяє гнучко визначати наступну конфігурацію обчислювальних ресурсів обслуговування, розміщувати мережеві функції з мінімальними витратами та децентралізовано контролювати показники якості.

Практичне значення одержаних результатів.

1. Розроблені механізми керування обчислювальними ресурсами обслуговування сервісів у гібридному середовищі були використані для обробки антарктичних даних у Національному антарктичному науковому центрі, проведені дослідження і отримані результати показали, що використання запропонованих рішень буде ефективним для вдосконалення існуючих телекомунікаційних мереж та при побудові нових перспективних гнучких добре масштабованих мереж оператора зв'язку із повним або частковим застосуванням принципу віртуалізації мережевих функцій.

2. Розроблені математичні моделі можуть бути використані при побудові блоку прийняття рішень щодо виділення обчислювальних ресурсів системи ЕРС для організації гібридного інформаційно-телекомунікаційного середовища, що гарантує якість обслуговування на заданому рівні під час обслуговування сервісів зі змінним трафіком, відповідні дослідження були проведені у лабораторіях компанії Lifecell.

3. Результати досліджень носять інтердисциплінарний характер, впроваджені у навчальний процес Інституту телекомунікаційних мереж КПІ

ім. Ігоря Сікорського. Проведено апробацію та виконано практичне тестування засобів керування перерозподілом ресурсів в лабораторіях телекомунікаційних компаній Huawei та Lifecell, результати моделювання запропонованих методів показали економію ресурсів обслуговування до 15%.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконувалась згідно із планами науково-дослідних робіт кафедри інформаційно-телекомунікаційних мереж КПІ ім. Ігоря Сікорського у рамках держбюджетних тем №2861-п «Технологія обробки сервісів з інтеграцією інформаційних ресурсів в системах підтримки операційної діяльності підприємств зв'язку» (державний реєстраційний номер – 0115U000217), №2117-п «Технологія побудови динамічних реєстрів електронних інформаційних ресурсів та засобів їх ефективної обробки у датацентрах гетерогенної структури» (державний реєстраційний номер – 0118U003522), госпдоговірної роботи: «Розробка системної інфраструктури єдиного інформаційного середовища даних антарктичних досліджень» та ініціативної роботи: «Технологія забезпечення якості обслуговування телекомунікаційних сервісів при частковому руйнуванні фізичного обладнання» (державний реєстраційний номер – 0117U001861).

Апробація результатів дисертації. Основні положення та результати дисертаційної роботи були представлені й одержали схвалення на міжнародній конференції «The Experience of Designing and Applications of CAD Systems in Microelectronics» (м. Львів, Україна, 2015 р.), міжнародній науково-технічній конференції «Проблеми телекомунікацій» (м. Київ, Україна, 2015, 2016, 2017, 2018 рр.), міжнародній науково-практичній конференції «Інфокомунікації – сучасність та майбутнє» (м. Одеса, Україна, 2015 р.), міжнародній конференції «Black Sea Conference on Communications and Networking» (Варна, Болгарія, 2016 р.), міжнародній конференції «Radio Electronics & Info Communications» (м. Київ, Україна, 2016 р.), міжнародній науково-практичній конференції «Problems of Infocommunications. Science and Technology» (м. Харків, Україна, 2017 р.).

Публікації. За результатами досліджень опубліковано 17 наукових праць, у тому числі 1 монографія, 6 статей у наукових фахових виданнях (з них 2 у виданнях України, які включені до міжнародних наукометричних баз), 9 тез доповідей в збірниках матеріалів конференцій, 1 стаття у інших виданнях.

Структура і обсяг дисертації. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел із 122 найменувань. Загальний обсяг роботи 173 сторінки, з яких 151 сторінка основного тексту, 15 сторінок використаних джерел та 7 сторінок додатків. Робота містить 78 рисунків, 3 таблиці.

РОЗДІЛ 1

КОНФІГУРАЦІЯ РЕСУРСІВ МЕРЕЖЕВИХ ФУНКЦІЙ В ГІБРИДНОМУ СЕРЕДОВИЩІ

1.1 Проблематика віртуалізації мережевих ресурсів

Сьогодні мобільні абоненти бажають залишатися на зв'язку в будь-якому місці, в будь-який час, і використовуючи будь-який пристрій. Це явище спонукає операторів мобільного зв'язку до побудови складних мережевих архітектур з включенням нових можливостей і розширень, якими важче керувати [1]. Два поняття знаходяться в центрі досліджень і розробок в даний момент, а саме Віртуалізація Мережевих Функцій (Network Functions Virtualization – NFV) і Програмно Конфігуровані Мережі (Software Defined Networking – SDN) [2].

Існуючі мобільні мережеві інфраструктури складаються з виділених мережевих вузлів, які зазвичай розміщуються в різних точках мережі, і кожному вузлу призначається надання певного набору функцій і сервісів. Така архітектура є негнучкою з точки зору впровадження нових послуг, здійснює неоптимальну маршрутизацію трафіку і неефективне використання ресурсів мережі. Крім того, масштабованість і економічність такої архітектури стає проблемою у світлі останніх прогнозів трафіку і очікувань користувачів [3].

З поширенням потужних мобільних пристроїв (наприклад, смартфонів, планшетних ПК і ноутбуків), а також зі зростаючою популярністю мобільних мультимедійних прикладних програм, очікуються підвищені вимоги до пропускної здатності. Згідно з [4] очікується, що загальний мобільний трафік даних зросте до 49 екзабайт на місяць до 2021 року, майже в сім разів більше в порівнянні з 2016. Мобільний трафік даних буде рости з середнім темпом річного зростання (CAGR) у 47 відсотків з 2016 до 2021 (рис. 1.1).

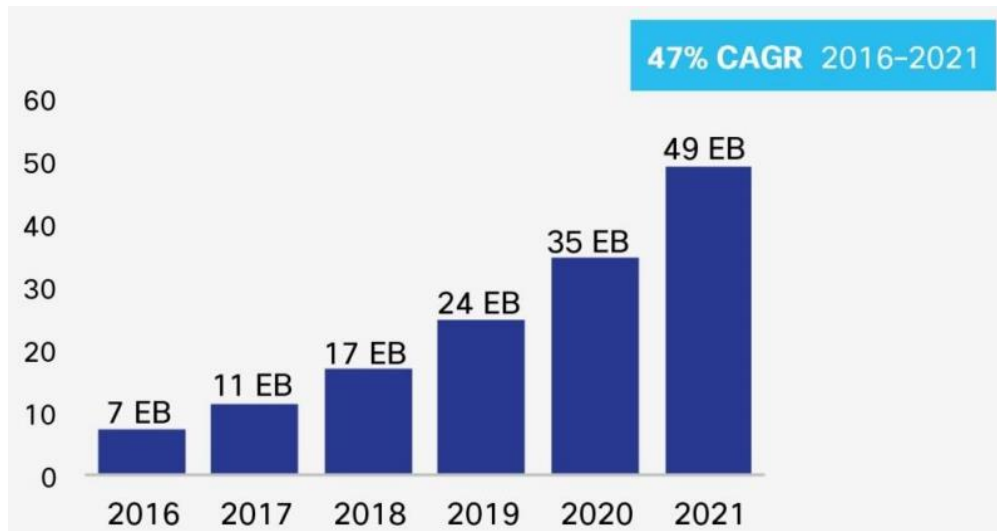


Рисунок 1.1 — Прогноз Cisco у 49 екзабайти на місяць трафіку мобільних даних до 2021 [4]

У мережах майбутніх поколінь, зв'язок machine-to-machine (M2M) буде поширюватися на широке коло речей, таких як транспортні засоби, побутова техніка, і інші подібні об'єкти, що призводить до Інтернету речей (Internet of Things – IoT) [5], комунікацій доповненої реальності, віртуальної реальності, а також інших передових технологій. Крім того, сервіси та прикладні програми будуть більш складними і диверсифікованими [6].

Як показано на рис. 1.2, транспортні засоби, будинки, персональні пристрої, роботи, датчики, і так далі всі будуть з'єднані безпроводними мережами. Тобто, буде досягнута автоматична і інтелектуальна система керування і контролю. Зростання числа пристроїв буде впливати на ринок IoT, який, за оцінками, має розмір у 19 трильйонів доларів США [7]. Число пристроїв, як очікується, досягне 50 мільярдів [8]. Крім того, повинні надаватися сервіси з багатим контентом, такі як потоковий перегляд фільмів у режимі реального часу, які потребують високої роздільної здатності, та телехірургія, що вимагають малого часу затримки [9].

Все безпроводно з'єднано Моніторинг/збір інформації & пристрої контролю		Розширення/збагачення безпроводових сервісів Багатий контент в реальному часі & гарантії безпеки	
Множинні персональні пристрої Взаємодія між багатьма пристроями	Транспортування (Машина/Автобус/Потяг) Розваги, навігація, інформація трафіку	Потокове відео 4K/8K відео роздільна здатність Відео на газетах Фонове відео	Новий тип терміналу Окуляри/Дотиковий Інтернет
Електроніка споживача Віддалена робота використовуючи персональний термінал	Годинник/прикраси/одяг Інтерфейс людини та сенсори догляду за здоров'ям	Охорона здоров'я Віддалена перевірка здоров'я та консультація	Освіта Віддалене навчання Будь-який урок будь-де/в будь-який час
Будинок Віддалений контроль обладнання Безпека дому	Сенсори Розумна електромережа Фермерство Автоматизація фабрик Погода/середовище	Хмарні обчислення Всі види сервісів підтримувані мобільною персональною хмарою	Система безпеки та рятування Попередження нещасних випадків Стійкість до лих

Рисунок 1.2 — Сервіси в епоху мереж майбутніх поколінь[9]

На рис. 1.3 [10] області А та В являють собою поточні мобільні послуги, що базуються на стільникових мережах. З іншого боку, область С представляє майбутні послуги на базі «Інтернету речей» (IoT), де використовується величезна кількість малопотужних терміналів, тоді як послуги в області D, такі як віддалена хірургія, будуть мати суворі вимоги до продуктивності (такі як низька затримка, висока надійність і невелика кількість терміналів), і будуть надаватися через майбутні базові мережі. Ця тенденція показує, що оператори мобільного зв'язку повинні відповідати вимогам сторонніх сервіс-провайдерів, таких як автомобільна промисловість, при збереженні операційних витрат на розумному рівні.

Останні нововведення в мобільних телекомунікаційних технологіях та мобільних терміналах стимулюють розповсюдження різних сервісів із широким діапазоном вимог щодо затримки, мобільності та надійності серед інших [11].

Надання послуг в телекомунікаційній галузі традиційно ґрунтується на тому, що мережеві оператори впроваджують фізичні пропріетарні пристрої та обладнання для кожної функції, яка є частиною певного сервісу. Крім того,

сервісні компоненти мають чіткі ланцюги і/або порядок, які повинні бути відображені в топології мережі і в локалізації сервісних елементів. Це, в поєднанні з вимогами до високої якості, стабільності і строгим дотриманням протоколу, привело до тривалих циклів продукту, дуже низької гнучкості обслуговування і сильної залежності від спеціалізованих апаратних засобів.

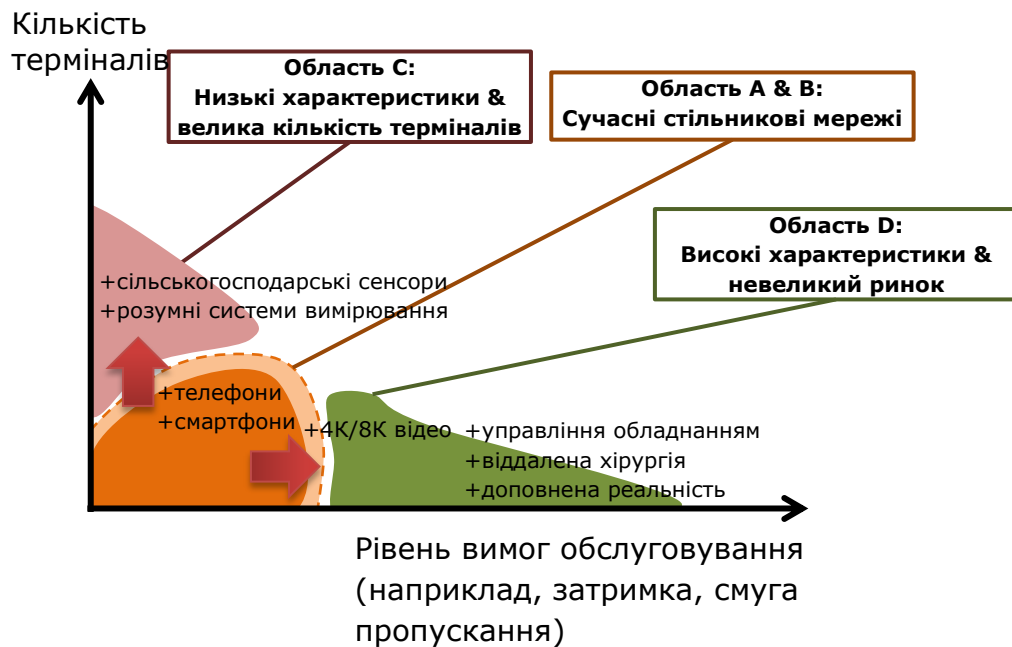


Рисунок 1.3 — Розширення сфери діяльності оператора [10]

Проте, вимоги користувачів до більш різноманітних і нових (короткоживучих) послуг з високими швидкостями передачі даних продовжують збільшуватися. Таким чином, телекомунікаційні сервіс-провайдери (Telecommunication Service Provider – TSP), повинні відповідним чином і постійно набувати, зберігати і експлуатувати нове фізичне обладнання. Це не тільки вимагає високих і швидко змінюваних навичок для техніків що експлуатують та управляють цим обладнанням, а й вимагає щільного розміщення мережевого обладнання, такого як базові станції. Все це призводить до високих капітальних і експлуатаційних витрат для TSP.

Більш того, навіть з цими високими вимогами абонентів, зростання капітальних і експлуатаційних витрат не може бути переведене в більш високу абонентську плату, так як TSP відомо, що через високу конкуренцію,

як між собою, так і від over-the-top послуг на каналах передачі даних, підвищення цін призводить лише до відтоку клієнтів. Тому TSP були змушені шукати шляхи побудови більш динамічних і сервіс-орієнтованих мереж з метою скорочення життєвих циклів продукції, операційних і капітальних витрат і підвищення оперативності обслуговування [12].

Значна залежність мереж від апаратного забезпечення та існування різних спеціалізованих апаратних пристроїв, таких як брандмауери, обладнання глибокої інспекції пакетів (DPI), і маршрутизаторів в мережевій інфраструктурі, посилює проблеми, що стоять перед провайдерами послуг мережі [13].

Як правило, оператори мобільного зв'язку справляються з підвищеними навантаженнями трафіку шляхом розширення/покращення загальної пропускної здатності мережі відповідним чином. Проте, це все більш і більш важко реалізувати за рахунок збільшення капітальних/операційних витрат (CAPEX/OPEX) в світлі низької рентабельності інвестицій (ROI). Окрім низького ROI, надмірне резервування ресурсів більше не вважається життєздатною стратегією, щоб задовольнити збільшення трафіку, так як відповідно до [14] до 80% обчислювальної потужності базових станцій і до половини потужності базової мережі є невикористаними. Це призводить до низького використання мережевих ресурсів, а також до високого рівня споживання енергії, що знижують економічну ефективність мережі для операторів мобільного зв'язку [3].

Принцип NFV спрямований на перетворення мережевих архітектур шляхом впровадження мережевих функцій в програмному забезпеченні, що може працювати на стандартній апаратній платформі. Крім того, він спрямований на перетворення традиційних мережевих операцій, оскільки програмне забезпечення може бути легко переміщене, або створено сутність в різних місцях без необхідності використовувати нове обладнання. NFV має багато переваг, від поліпшення операційної ефективності і зниження енергоспоживання до коротших інтервалів розгортання/оновлення і майже

оптимального використання мережевих ресурсів, оскільки будівельні блоки можуть виділятися і перерозподілятися під час виконання в залежності від вимог [15, 16].

Хмарна мобільна мережа це [17]: мережа мобільного зв'язку + децентралізовані обчислення + смарт зберігання, що пропонується як один сервіс, який надається за вимогою, еластично та з оплатою Pay-As-You-Go.

Ідеєю хмарної мобільної мережі є розширення концепції хмарних обчислень за межі центрів обробки даних у напрямку до мобільного кінцевого користувача (як показано на рис. 1.4).

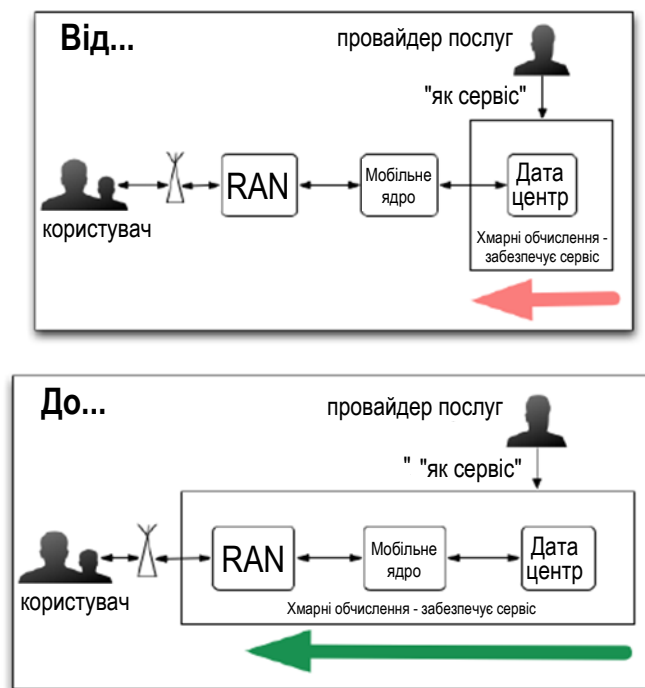


Рисунок 1.4 — Розширення концепції хмарних обчислень за дата центри до мобільного кінцевого користувача [17]

NFV прокладає шлях до ряду відмінностей в способах реалізації надання мережевого сервісу в порівнянні з існуючою практикою. Таким чином, ці відмінності можна охарактеризувати наступним чином [12]:

1. Відв'язка програмного забезпечення від апаратного. Оскільки елемент мережі більше не є об'єднанням інтегрованих апаратних і програмних сутностей, еволюція обох є незалежною один від одного. Це

дозволяє мати окремі терміни розробки і технічного обслуговування програмного і апаратного забезпечення.

2. Гнучке розгортання мережевих функцій. Відрив програмного забезпечення від апаратного допомагає перерозподілити і спільно використовувати ресурси інфраструктури, таким чином, разом, апаратне і програмне забезпечення, може виконувати різні функції в різний час. Це допомагає мережевим операторам розгорнути нові мережеві сервіси швидше по тій же фізичній платформі. Таким чином, компоненти можуть бути створені в будь-якому NFV-сумісному пристрої в мережі і їх з'єднання можуть бути встановлені на гнучкій основі.

3. Динамічне масштабування. Розділення функціональності мережевої функції на створювані програмні компоненти забезпечує більшу гнучкість масштабування реальної продуктивності віртуальної мережевої функції (Virtual Network Function – VNF) більш динамічно і з більшою деталізацією, наприклад, відповідно до фактичного трафіку, для якого оператор мережі повинен надавати ємність.

Таким чином, метою NFV є трансформація способу, яким оператори мереж і провайдери мережевих послуг будують, керують та розгортають мережеву інфраструктуру, завдяки еволюції технологій віртуалізації. Це перетворення виконується за допомогою консолідації різних типів віртуалізованих мережевих функцій в стандартних комп'ютерах загального призначення (серверах, пристроях зберігання даних і т.д.), які можуть бути розташовані в дата центрах, мережевих вузлах і близько до приміщень кінцевого користувача (див. рис. 1.5) [18].

З масовим розповсюдженням LTE оператори стикаються з сигналізацією, яка загрожує завалити традиційні базові мережі. Оскільки абоненти стають дедалі активнішими в соціальних мережах і одночасно залучають все більше прикладних програм, смартфони можуть створювати величезні обсяги сигналізації, коли вони взаємодіють із мережею.

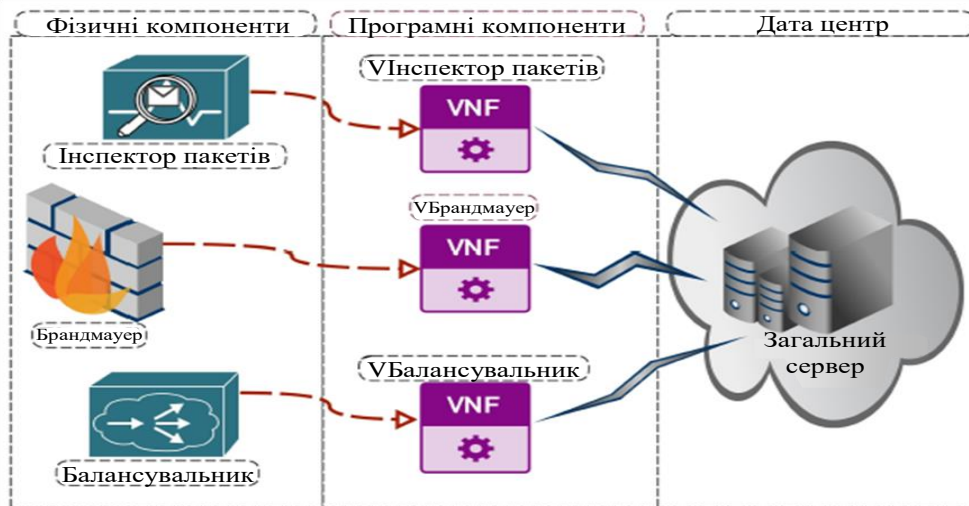


Рисунок 1.5 — Середовище NFV

Проникнення мобільного широкосмугового зв'язку буде збільшуватися, і як наслідок, зростання трафіку сигналізації значно перевищить відповідний приріст трафіку даних. Компанія Nokia Siemens Networks [19] прогнозує, що зростання трафіку сигнальних повідомлень буде на 50% швидше, ніж зростання трафіку даних протягом найближчих кількох років. Проте збільшення використання смартфонів є лише одним із факторів вибуху сигнального трафіку.

Спонукальні чинники сигнального трафіку (рис. 1.6):

- оператори все більше покладаються на поінформованість про обслуговування, щоб забезпечити кращий досвід для користувачів смартфонів; у поєднанні з диференційованою тарифікацією, результат у збільшенні сигналізації становить до 30% (див. рис. 1.7);
- зростаюча кількість підключених до Інтернету мобільних пристроїв міжмашинної взаємодії та прикладних програм із високими вимогами мобільності призведе до значної сигналізації;
- Voice over LTE вимагає масштабованої пропускної здатності сигналізації та диференційованої якості сприйняття сервісів в режимі реального часу;

- виникаючі тенденції та нові бізнес-моделі вимагають масштабованості площин користувача та управління;
- зростаюче використання онлайн-маркетингу, хмарного сховища, контент-бізнесу та прикладних програм;
- збільшення кількості клієнтів, які мають декілька персональних пристроїв;
- конвергенція таких галузей, як медіа, соціальні мережі, охорона здоров'я, енергетика та екологічні послуги тощо.

Еволюція мережі	Еволюція розумних пристроїв	Еволюція додатків і сервісів
Абоненти мереж LTE генерують утричі більше сигналізації у EPS	Двозначний ріст у впровадженні смартфонів в поєднанні з реалізаціями операційних систем конкретних постачальників призводить до збільшення сигналізації до 50%.	Сигналізація OverTheTop-гравців, включаючи багатий на відео контент, онлайн-ігри та мобільні додатки, знаходиться поза контролем операторів.

Рисунок 1.6 — Різноманітні причини зростання трафіку сигналізації

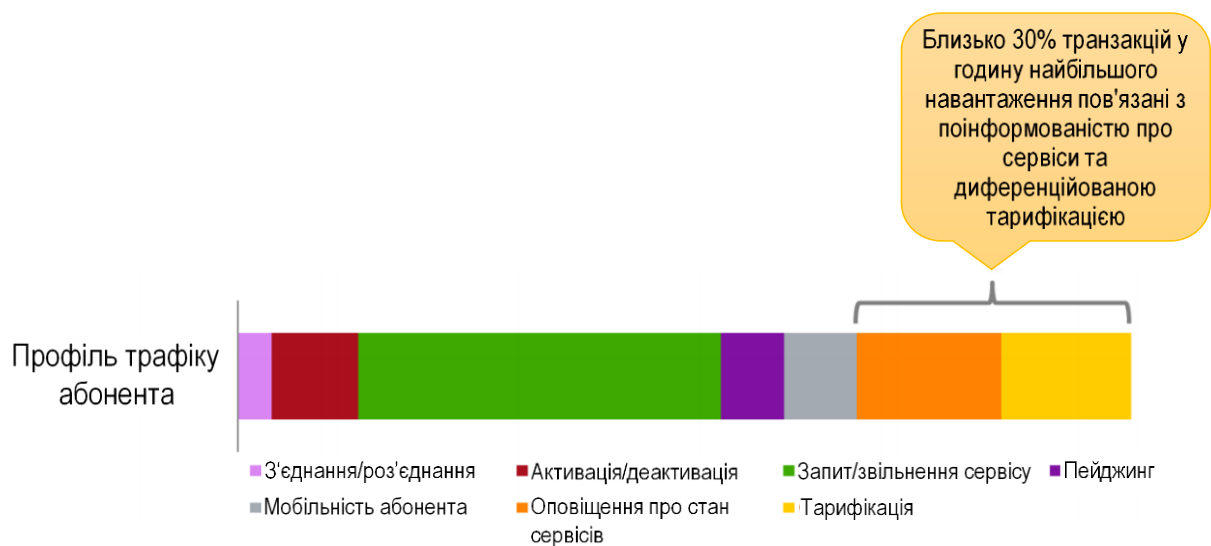


Рисунок 1.7 — Поінформованість щодо сервісів та диференціація тарифікації можуть додати додаткові 30% до загального навантаження сигналізації у мережах [19]

У мережі 3G, контролер радіомережі (Radio Network Controller – RNC), знаходиться між базовою станцією і елементами базової мережі, ефективно

захищаючи базову мережу від великої кількості сигналізації, що генерується мережею радіодоступу для керування мобільністю. На відміну від цього, LTE використовує плоску архітектуру, яка усуває RNC. Оскільки базова мережа підключена безпосередньо до базових станцій LTE, то вона повинна обробляти весь трафік сигналізації (див. рис. 1.8).

Середня вимога сигналізації на одного абонента до 42% вище у LTE порівнюючи з HSPA. Зрозуміло, що сигналізація стала вирішальним фактором при визначенні розмірів базової мережі [19].

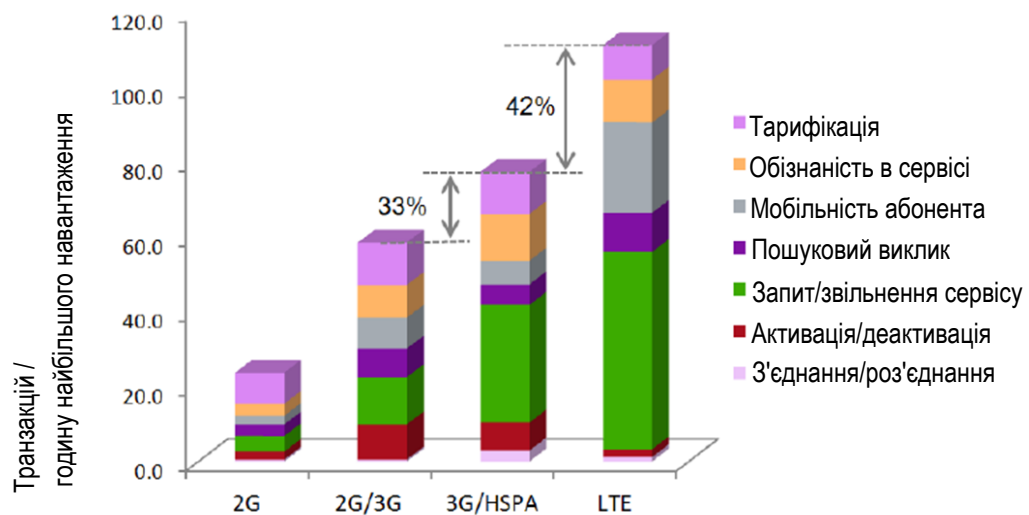


Рисунок 1.8 — Середні вимоги сигналізації на одного абонента [19]

Еволюційних шлях архітектури базової мобільної мережі представлено на рис. 1.9 [20].

Однією з основних переваг NFV є обіцянка підвищити ефективність використання енергії в результаті консолідації ресурсів, а також їх більш динамічного використання. У [21] визначили, що частина мобільної мережі з найвищими перспективами використання енергії є Evolved Packet Core (EPC), де віртуалізація функцій призводить до скорочення на 22% споживання енергії та підвищення на 32% в області енергоефективності (рис. 1.10).

У EPC, яка є останньою архітектурою базової мережі для системи стільникового зв'язку, до прикладів мережевих функцій відносяться MME, S/P-GW, і т.д. HSS і PCRF є іншими мережевими функціями 3GPP, які

необхідні в архітектурі для надання сервісу з кінця в кінець. Аналогічним чином, онлайн і оффлайн системи тарифікації (OCS і OFCS) представляють собою системи, які фіксують записи тарифікації в рамках управління сесією.

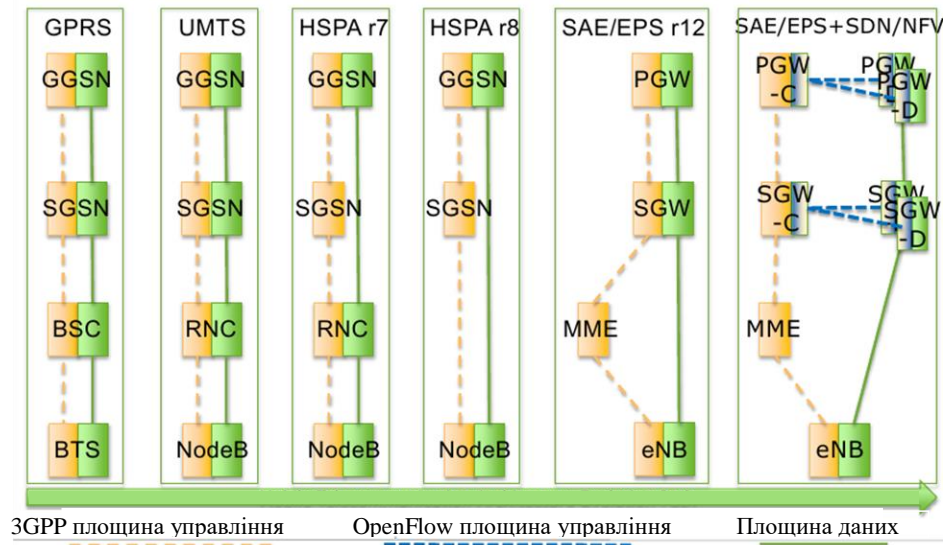


Рисунок 1.9 — Еволюційних шлях архітектури базової мобільної мережі

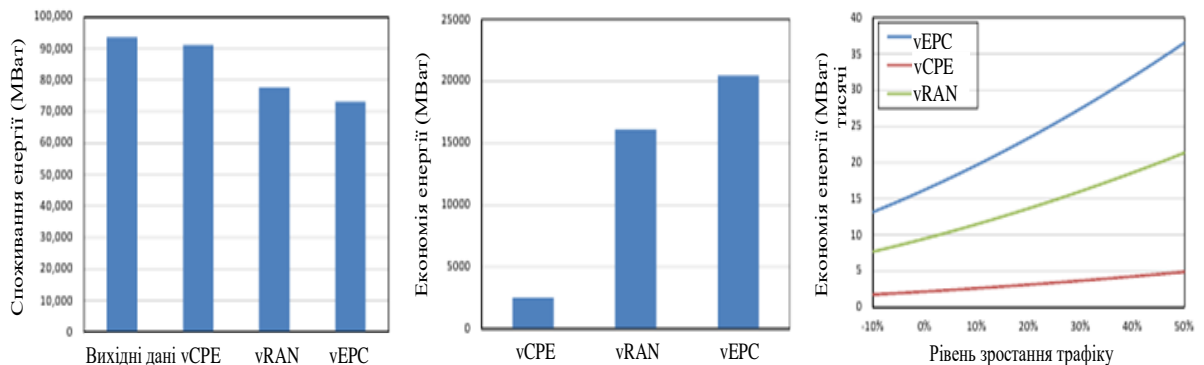


Рисунок 1.10 — Енергоспоживання у віртуалізованих мережах а) загальне енергоспоживання б) економія енергії в) варіація економії енергії [21]

Таким чином, одним із варіантів застосування NFV є віртуалізація EPC [22].

З часом EPC стане загальною базовою мережею для послуг з комутацією пакетів у мережах 2G та 3G, а також потенційно для Wi-Fi. EPC є центральним елементом архітектури мобільної мережі та важливим для надання сервісів кінцевим користувачам. Вона розташована між мережею

радіодоступу (RAN), IP-мережею, зовнішніми сервісами та низкою інших мережевих елементів, що використовуються для автентифікації, захисту та надання послуг кінцевим користувачам. Це стратегічна точка контролю за наданням послуг мобільної передачі даних, як показано нижче на рис. 1.11 [23].



Рисунок 1.11 — EPC як важлива складова мобільної мережі [23]

EPC є однією з ключових областей, яку оператори та вендори визначили як хороший кандидат для NFV. Наприклад, визначається як життєздатний і привабливий випадок використання в специфікаціях робочої групи ETSI.

У [23] також виявлено великий інтерес до віртуальних мобільних базових мереж серед операторів через ряд власних дослідницьких проєктів та залучень операторів, здійснених останнім часом. Є кілька причин, чому EPC вважається гарним кандидатом для віртуалізації. Комерційно, це нова та така, що розширюється, інвестиція, яке все ще має тривалий час життя попереду, тому узгодження інвестицій в мережу з майбутніми технологічними тенденціями є логічним. Технічно, природа EPC та порівняно скромні обсяги трафіку в мобільних мережах підходять для віртуалізації, хоча це дещо залежить від конкретної моделі застосування та розгортання.

Враховуючи важливу роль EPC у забезпеченні гарантованого надання послуг з найкращою якістю сприйняття (Quality of Experience – QoE) та оскільки мобільна мережа зазнає трансформацію, а механізми надання

послуг удосконалюються, операторам мобільних мереж необхідно розуміти, визначати та вимірювати економічні вигоди від віртуалізації. Деякі з них [24]:

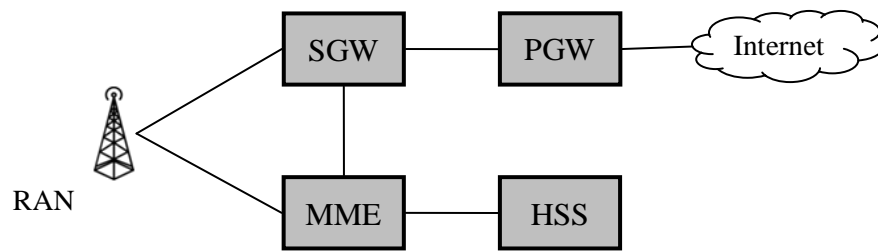
- показано, що віртуалізація всього комплексу функцій EPC дає змогу інфраструктурі мобільної мережі працювати на більш високому рівні використання до 87%, що безпосередньо призводить до підвищення ефективності експлуатаційних витрат до 25%;

- гнучкий рівень оркестровки може скоротити час виходу на ринок сервісу (Time To Market – TTM) шляхом спрощення критичних засобів організації, розгортання, автоматизації, взаємодії, настройки та оптимізації, а також всебічної доступності мережевих функцій EPC. Зокрема, в дослідженні [24] було виявлено зниження TTM на 67% для запуску нових сервісів, таких як MVNO, приватна мобільна мережа, випадки використання IoT та VoLTE.

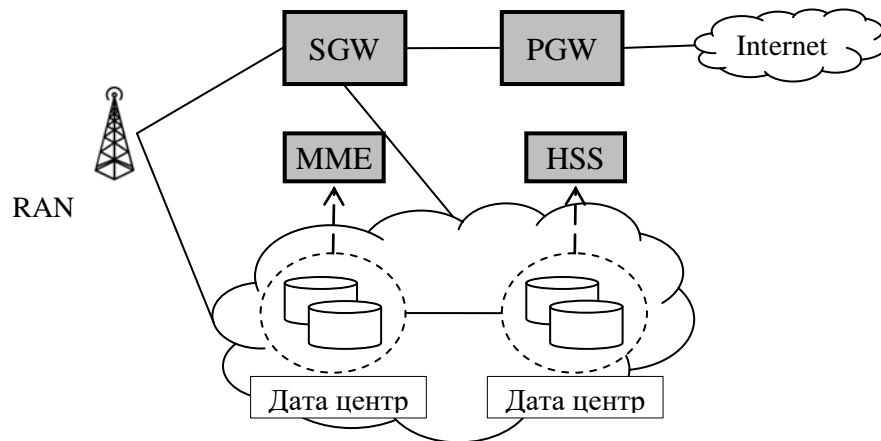
На практиці, більш дорогі спеціалізовані апаратні засоби часто працюють швидше і ефективніше ніж віртуалізовані сутності, навіть хоча останні є більш гнучкими. Оскільки спеціалізовані апаратні засоби на даний час широко використовуються, цілком імовірно, що гібридні сценарії розгортання стануть поширеними, коли частина сервісів надається фізичним обладнанням (рис. 1.12). У NFV мережах, набір ланцюгів сервісів повинен розташовуватись на фізичних вузлах мережі. Ланцюг сервісів – це набір з одного або декількох сервісів або віртуальних машин, які з'єднані разом для забезпечення певної функціональності, і можуть бути представлені у вигляді графа, що містить сервіси і мережеві вимоги між цими сервісами. У гібридному мережевому середовищі, ланцюги сервісів можуть розміщуватись або з використанням фізичного обладнання, або з використанням віртуалізованих сутностей. Успіх цього підходу залежить від наявності та продуктивності алгоритмів, що визначають де і як ці структурні блоки створюються [16].

Поточна архітектура EPC показана на рис. 1.12 (а). Архітектура EPC є мережею IP з комутацією пакетів, яка складається з різних вузлів: Evolved

NodeB (eNB) вузла для радіодоступу LTE, вузла управління мобільністю (Mobility Management Entity – MME) – вузол для управління мобільності терміналу, вузла домашньої абонентської системи (Home Subscriber System – HSS), вузла бази даних інформації про користувачів, вузла обслуговуючого шлюза (Serving Gateway – SGW), вузла якірної точки мобільності для управління терміналами, а також вузла шлюза пакетної мережі передачі даних (Packet Data Network Gateway – PGW), який служить шлюзом між терміналами і зовнішніми мережами, такими як Інтернет [9].



(а) Типова мережа LTE/EPC



(б) Часткова віртуалізація LTE/EPC

Рисунок 1.12 — Мобільна базова мережа

1.2 Огляд архітектури LTE/EPC

На рис. 1.13 показана типова архітектура LTE/EPC, яка показує мережеві елементи, які з'єднані між собою через чітко визначені інтерфейси, і кожному елементу мережі присвоюється виділений набір спеціалізованих функцій/сервісів [3].

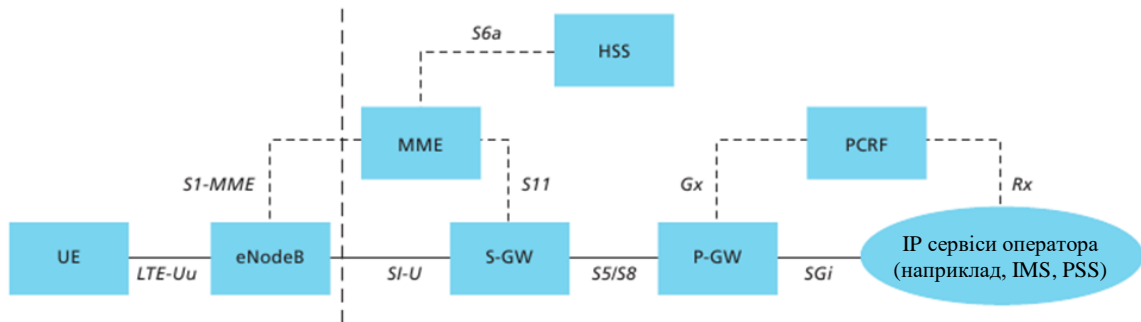


Рисунок 1.13 — Мережні елементи EPS [25]

Evolved Packet System (EPS) надає користувачеві IP підключення до Packet Data Network (PDN) для доступу в Інтернет, а також для запуску сервісів, таких як передача голосу по IP (Voice over IP – VoIP). EPS канал (bearer), як правило, асоціюється з якістю обслуговування (Quality of Service – QoS). Множинні канали можуть бути встановлені для користувача, для того, щоб забезпечити різні потоки QoS, а також підключення до різних PDN. Наприклад, користувач може здійснювати голосовий виклик (VoIP), в той же час виконуючи перегляд веб-сторінок або здійснюючи FTP-завантаження. VoIP канал забезпечить необхідну QoS для голосового виклику, в той час як канал з «найкращими зусиллями» буде цілком відповідним для перегляду веб-сторінок або FTP-сесії. Мережа повинна також забезпечити достатню безпеку і конфіденційність для користувача і захист мережі від несанкціонованого використання.

Це досягається за допомогою декількох мережевих елементів EPS, які мають різні ролі. На рис. 1.13 показана загальна архітектура мережі, включаючи мережеві елементи і стандартизовані інтерфейси. На високому рівні мережа складається з базової мережі (Core Network) (EPC) і мережі доступу E-UTRAN. У той час як базова мережа складається з безлічі логічних вузлів, мережа доступу складається з по суті тільки одного вузла, базової станції Evolved NodeB (eNodeB), який підключається до абонентського обладнання (UE). Кожен з цих елементів мережі з'єднаний за допомогою інтерфейсів, які стандартизовані для того, щоб дозволити функціональну

сумісність різних постачальників. Це дає операторам мережі можливість використовувати різні мережеві елементи від різних постачальників. Насправді, в їх фізичних реалізаціях мережеві оператори можуть вибрати розділити або об'єднати ці логічні мережеві елементи в залежності від комерційних міркувань. Функціональний розподіл між EPC і E-UTRAN показан на рис. 1.14.

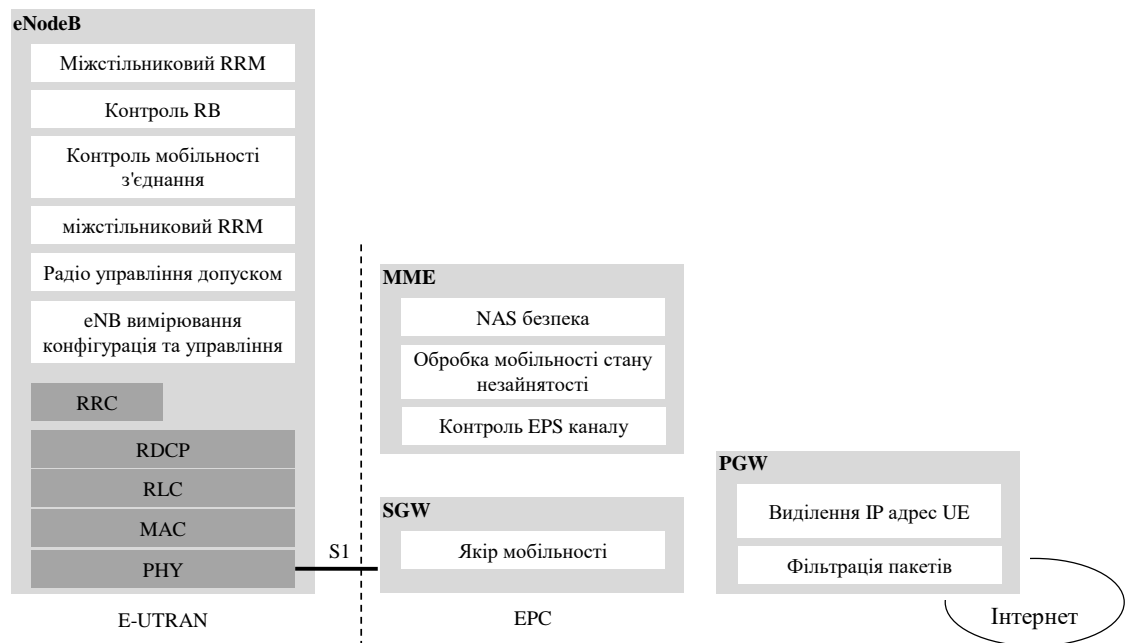


Рисунок 1.14 — Функціональний розподіл між E-UTRAN та EPC [25]

Базова мережа (яка називається EPC в LTE) відповідальна за загальний контроль над UE і встановлення каналів. Так як EPS забезпечує тільки шлях каналу певної QoS, управління мультимедійними додатками, такими як VoIP, забезпечується IP мультимедійною підсистемою (IP Multimedia Subsystem – IMS), яка, вважається, розташовується поза самої EPS.

Логічні вузли базової мережі показані на рис. 1.14 і більш докладно описані нижче [25]:

- PCRF – PCRF відповідає за прийняття рішень контролю політики, а також за управління тарифікацією на основі потоку в функції впровадження політики управління (Policy Control Enforcement Function – PCEF), що знаходиться в PGW. PCRF забезпечує авторизацію QoS (ідентифікатор класу

QoS (QCI) і швидкість передачі бітів), що вирішує, як певний потік даних буде оброблятися в PCEF і гарантує, що це відповідає профілю підписки користувача.

– HSS – домашній абонентський сервер містить дані підписки користувачів SAE, такі як EPS-підписний профіль QoS і будь-які обмеження доступу для роумінгу. Він також містить інформацію про PDN, до яких користувач може підключитися. Це може бути у вигляді імені точки доступу (Access Point Name – APN) (яке являє собою позначку згідно з угодами про присвоєння імен DNS, що описує точку доступу для PDN) або адреси PDN (зазначаючи IP адресу/адреси підписки). Крім того, HSS має динамічну інформацію, таку як ідентифікатор MME, до яких користувач в даний момент підключений або зареєстрований. HSS також може інтегрувати центр аутентифікації (Authentication Center – AUC), який генерує вектори для ключів аутентифікації і безпеки.

– PGW – PDN Шлюз відповідає за розподіл IP адрес для UE, а також за забезпечення дотримання QoS і тарифікацію на основі потоку відповідно до правил з PCRF. Він відповідає за фільтрацію IP пакетів користувача низхідної лінії зв'язку в різні канали на основі QoS. Це виконується на основі шаблонів потоку трафіку (Traffic Flow Templates – TFTs). P-GW виконує впровадження QoS для каналів з гарантованою швидкістю передачі даних (Guaranteed Bit Rate – GBR). Він також служить як якор мобільності для взаємодії з не-3GPP технологіями, такими як мережі CDMA2000 і WiMAX.

– SGW – всі IP пакети користувача передаються через Обслуговуючий Шлюз, який служить локальним якорем мобільності для каналів передачі даних, коли UE переміщається між eNodeB. Він також зберігає інформацію про канали, коли UE знаходиться в стані очікування (відомий як «EPS управління з'єднанням – IDLE» [ECM-IDLE]) і тимчасово буферизує даних низхідної лінії в той час як MME ініціює пошуковиц виклик UE для повторного встановлення каналу. Крім того, S-GW виконує деякі

адміністративні функції в гостьовій мережі, такі як збір інформації для тарифікації (наприклад, обсяг даних, переданих або отриманих від користувача) і законне перехоплення. Він також служить в якості якоря мобільності для взаємодії з іншими технологіями 3GPP, такими як GPRS і UMTS.

- MME – Mobility Management Entity (MME) є вузлом управління, який обробляє сигналізацію між UE і базовою мережею. Протоколи, що знаходяться між UE і базовою мережею відомі як протоколи Non Access Stratum (NAS).

Основні функції, підтримувані MME можуть бути класифіковані як:

- функції, пов'язані з управлінням каналу – включає в себе встановлення, підтримку і звільнення каналів і обробляється рівнем управління сесією в протоколі NAS;

- функції, пов'язані з управлінням з'єднанням – включає в себе встановлення з'єднання і безпеку між мережею і UE і обробляється рівнем з'єднання або управління мобільністю в рівні протоколу NAS.

Віртуалізовані мережеві функції (VNF), наприклад, S/P-GW, MME, можуть масштабуватися незалежно один від одного відповідно до їх конкретних потреб в ресурсах, наприклад може виникнути ситуація, коли необхідно збільшити ресурси площини користувача, не впливаючи на площину управління, і навпаки. Крім того, VNF, які мають справу з площиною даних може знадобитися різна кількість ресурсів NFV інфраструктури, ніж тим VNF, що займаються тільки сигналізацією.

Різні сценарії можуть існувати, де, наприклад, весь EPC віртуалізовано в одній точці присутності інфраструктури NFV (NFVI) або тільки деякі мережеві функції віртуалізовано.

Для досягнення безперервності обслуговування і доступності послуг бажаних оператором, надійність повинна бути забезпечена і в площині управління, і в площині даних. Оскільки віртуалізація забезпечує відв'язку

мережевої функції від нижчезрештованого обладнання, проектування нових схем відмовостійкості стає можливим за рахунок використання портативності екземплярів VNF. На рис. 1.15 [22] показано віртуалізовані EPC і IMS, де віртуалізовані S/P-GW і функції IMS оброблюють PDN з'єднання і IMS сесії відповідно. Коли динамічне переміщення цих екземплярів VNF виконується через перевантаження віртуальної машини або відмову в автоматичному режимі або на вимогу, переміщення керованих сесій і/або з'єднань повинні бути оброблені відповідним чином для досягнення бажаних оператором безперервності обслуговування і доступності послуг.

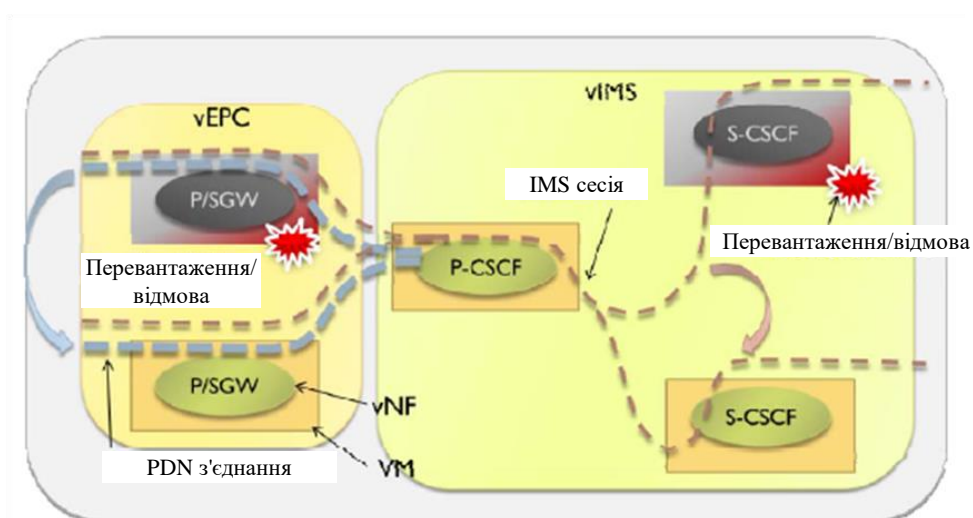


Рисунок 1.15 — Переміщення VNF [22]

Нижче перераховані проблеми високого рівня, які повинні бути прийняті до уваги при визначенні конкретних рішень для цього випадку використання [22]:

1. Масштабування ресурсів: розширення і згортання мережевих ресурсів віртуалізованих мережевих функцій.
2. Обізнаність в сервісах: сервіс-обізнаний розподіл ресурсів для мережевих функцій.
3. Прозорість віртуалізації для сервісів: сервіси використовуючи мережеву функцію не повинні знати, чи це віртуальна функція або не-віртуалізована.

4. Прозорість віртуалізації для мережевого контролю і управління: площа мережевого контролю та управління не повинна бути обізнаною в тому, чи функція віртуалізована чи ні.

5. Управління станом: управління станом мережі і мережевої функції під час перенесення мережевої функції, реплікації і масштабування ресурсів.

6. Моніторинг/виявлення помилок/діагностика/відновлення: відповідний механізм для моніторингу/виявлення помилок/діагностики/відновлення всіх компонентів і їх станів після віртуалізації, наприклад, екземплярів VNF, апаратного забезпечення, гіпервізора.

7. Доступність сервісу: досягнення такого ж рівня доступності сервісів для віртуалізованої мобільної опорної мережі з кінця в кінець, як в не віртуалізованих мережах з меншими витратами.

8. Механізм поділу управління трафіком: ідентифікація/поділ трафіку даних і управління для не-віртуалізованих і віртуалізованих мобільних базових мереж.

9. Вплив на функції: зведення до мінімуму впливу на існуючі не-віртуалізовані мережеві функції та підтримка мережевих операційних систем.

EPaaS можна розглядати як хмарну реалізацію 3GPP EPC архітектури, що може надаватися на вимогу [19], за допомогою архітектурної платформи, описаної в цій роботі.

1.3 Якість обслуговування та EPS канали

Розглянемо вимоги до якості обслуговування в мережах LTE [25].

Канали можуть бути класифіковані на дві категорії, основуючись на характері QoS, які вони забезпечують:

1. Канали з мінімальною гарантованою швидкістю передачі даних (Guaranteed Bit Rate – GBR).

2. Не-GBR канали, які не гарантують будь-якої конкретної швидкості передачі даних.

Кожен канал має пов'язаний з ним ідентифікатор класу якості обслуговування (QoS Class Identifier – QCI), а також пріоритет розподілу і утримання (Allocation and Retention Priority – ARP). Кожен QCI характеризується пріоритетом, граничним допустимим значенням затримки пакетів і прийнятним рівнем втрат пакетів, набір стандартизованих QCI і їх характеристик можна знайти в розділі 6.1.7 в [26]. Пріоритет і гранична затримка пакетів (і до певної міри прийнятний коефіцієнт втрати пакетів) з мітки QCI визначають конфігурацію режиму Radio Link Control (RLC) (дивіться розділ 4.3.1 з [27]) і як планувальник в MAC обробляє пакети, що передаються через канал (наприклад, з точки зору політики планування, політики управління чергами та політики формування швидкості). ARP каналу використовується для контролю допустимості викликів – тобто, щоб вирішити, чи слід встановлюватися запитуваний канал в разі радіо перевантаження.

1.4 Віртуалізація Мережевих Функцій (NFV)

1.4.1 Високорівнева платформа NFV

NFV передбачає реалізацію мережевих функцій як сутностей програмного забезпечення, які працюють на Інфраструктурі NFV (NFVI). На рис. 1.16 [28] показано платформу NFV високого рівня. Таким чином, три основні робочі області визначені в NFV [29]:

- Віртуалізована Мережева Функція (VNF), як програмна реалізація функції мережі, яка здатна працювати на NFVI;
- Інфраструктура NFV (NFVI), включаючи різноманітність фізичних ресурсів і як вони можуть бути віртуалізовані. NFVI підтримує виконання VNF;

– Управління і Оркестровка NFV (NFV MANO), що охоплює оркестровку і управління життєвим циклом фізичних і/або програмних ресурсів, що підтримують віртуалізацію інфраструктури, та управління життєвим циклом VNF. Управління і Оркестровка NFV фокусується на всіх специфічних для віртуалізації завданнях управління необхідних в платформі NFV.

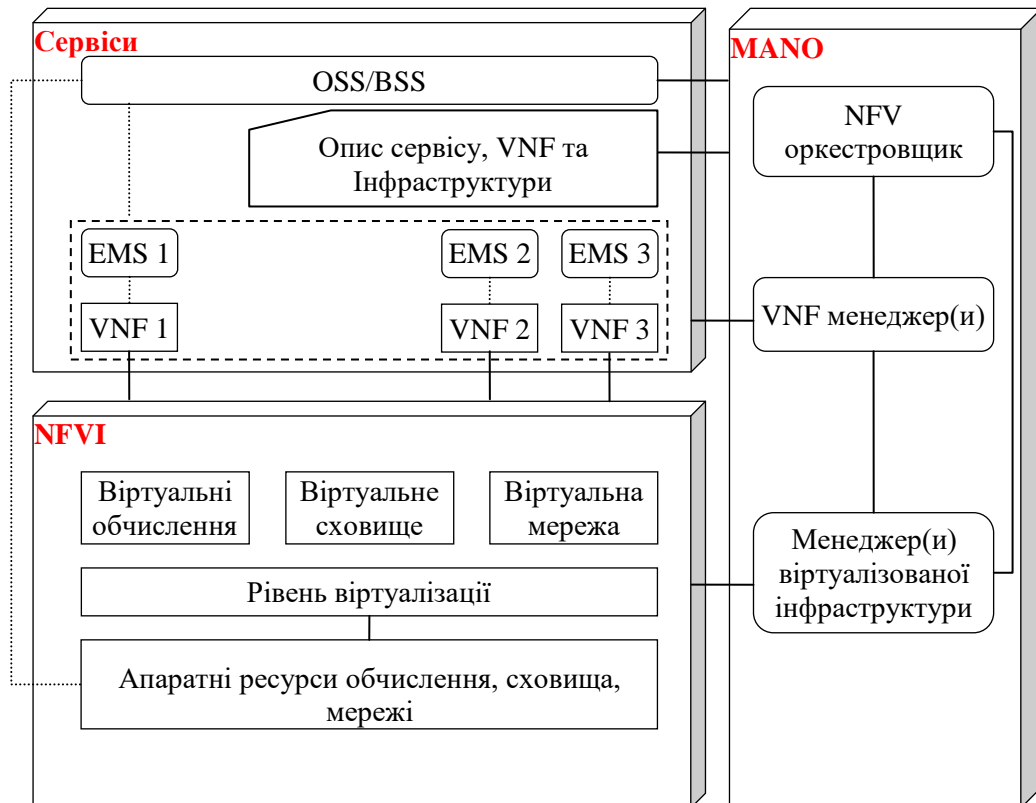


Рисунок 1.16 Високорівнева платформа NFV [28]

На рис. 1.17 [30] MANO – це виділений блок праворуч, що складається з трьох менеджерів і групи репозиторіїв (4):

1. Virtualized Infrastructure Manager (VIM).
2. VNF Manager (VNFM).
3. NFV Orchestrator (VNFO).
4. Репозиторії NS Catalog, VNF Catalog, NFV Instances і NFVI Resources.

На додаток до чотирьох блоків всередині MANO, на схемі присутні два блоки – традиційні Element Management System (EMS) і OSS/BSS. Хоча

останні два блоки не є частиною MANO, вони обмінюються з нею інформацією.

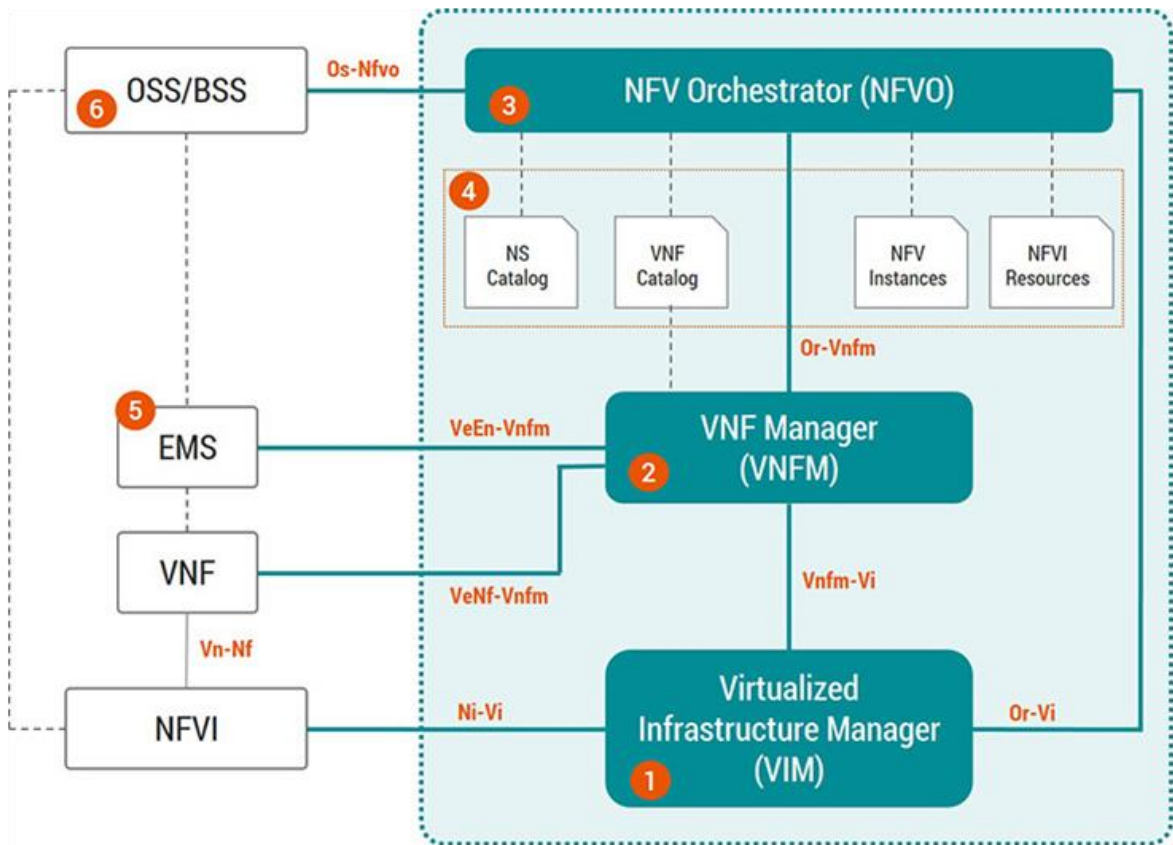


Рисунок 1.17 — ETSI-NFV архітектура [30]

Virtualized Network Function (VNF) – це сама віртуальна функція (наприклад, віртуальний DPI, Firewall, Router і ін.).

NFV Infrastructure (NFVI) визначає інфраструктуру – фізичні (сервери, диск та ін.), віртуальні ресурси (віртуальні машини) і гіпервізор.

Функції, які виконує VIM:

- управління життєвим циклом ресурсів в домені NFVI, включаючи створення, управління станом і видалення віртуальних машин;
- зберігання бази віртуальних машин і асоційованих з ними фізичних ресурсів;
- моніторинг продуктивності, виявлення відмов обладнання, програмного забезпечення та віртуальних ресурсів;

- надає інтерфейс (Northbound API) іншим системам управління для надання інформації про фізичні та віртуальні ресурси.

VNFM управляє віртуальної мережевої функцією (VNF). Він як VIM для NFVI, тільки для VNF.

Функції, які він виконує:

- управління життєвим циклом VNF – створення, управління станом і видалення VNF Instances (самі VNF встановлені на віртуальних машинах, які знаходяться під управлінням VIM);
- конфігурація VNF, моніторинг і статистика, метрики продуктивності і управління політиками безпеки;
- управління розміром VNF – збільшення/зменшення ресурсів CPU, пам'ять, диск.

NFV Orchestrator (NFVO) виконує наступні функції:

1. Оркестрація ресурсів. NFVO управляє виділенням і резервуванням ресурсів і дає дозволи на їх використання або в рамках одного або декількох доменів NFVI. Управління здійснюється через відповідні VIM за допомогою API, а не за допомогою прямої взаємодії з ресурсами NFVI.

2. Оркестрація сервісів. NFVO відповідає за створення призначених для користувача сервісів з декількох VNF (які можуть управлятися різними VNF менеджерами). Оркестрація сервісів відбувається наступним чином:

- NFVO створює сервіс за допомогою взаємодії з відповідним VNFM, без необхідності прямої взаємодії з VNF;
- NFVO може звертатися до VNFM для отримання будь-якої інформації про VNF;
- зберігає топологію створених ланцюжків сервісів (англ. VNF Forwarding Graphs або Service Chains).

NFVO «склеює» воєдино інші функції MANO і надає уніфікований інтерфейс для управління ресурсами і сервісами в рамках мультивендорної платформи.

Репозиторії в NFV MANO зберігають різну інформацію, в залежності від їх типу.

VNF Catalog – репозиторій всіх використовуваних VNF дескрипторів (VNFD). VNF Descriptor – шаблон, який визначає параметри VNF і вимоги до впровадження та експлуатації. Використовується VNF менеджером в процесі створення VNF і управління життєвим циклом віртуальної мережевої функції. Інформація, яку надає VNF дескриптор, також використовується NFVO для оркестрації мережевих сервісів і інфраструктурних ресурсів.

Network Services (NS) Catalog – список використовуваних для користувача мережевих сервісів. Шаблон мережевого сервісу надає опис взаємодії VNF між собою (Наприклад, характеристики віртуального каналу).

NFV Instances – список даних про прив'язку користувальницьких мережевих сервісів і відповідних VNF. Наприклад, два різних клієнта придбали віртуальний маршрутизатор. Щоб відрізнити віртуальний маршрутизатор клієнта 1 від віртуального маршрутизатора клієнта 2, необхідний список з прив'язкою ID маршрутизаторів до конкретного клієнта.

NFVI Resources – репозиторій для зберігання використовуваних ресурсів. Необхідний для створення нових VNF сервісів або масштабування існуючих.

EMS відповідає за конфігурацію VNF, моніторинг і збір статистики, метрики продуктивності і управління політиками безпеки. Хоча VNFM робить те ж саме, однак, EMS виконує ці функції за допомогою пропрієтарних/закритих інтерфейсів взаємодії з VNF на відміну від VNFM. Потрібен відкритий «інтерфейс» взаємодії (Ve-Vnfm-em), в MANO званий Reference Point.

Взаємодія зовнішньої EMS з VNFM може знадобитися для отримання інформації щодо віртуальних ресурсів, асоційованих з даною мережевою функцією (VNF).

OSS/BSS – IT-системи оператора. Природно, NFV повинна працювати в тісній взаємодії з цими системами.

В принципі можна було розширити функціональність існуючих OSS/BSS систем і «навчити» їх управляти VNF і NFVI безпосередньо. Але в більшості випадків все впиралося б в реалізацію від конкретного виробника. NFV позиціонується як відкрита платформа, і управління її елементів повинно здійснюватися через відкриті інтерфейси (як в MANO).

Проте, існуючі OSS/BSS можуть додати бізнес-переваг до NFV MANO шляхом пропозиції додаткових функцій, які не підтримуються конкретною реалізацією архітектури. Для цього в архітектурі передбачений ще один Reference Point (Or-Ma-NFVO), що описує взаємодію MANO з OSS/BSS системами.

1.4.2 Функції та сервіси віртуальної мережі

Мережева функція є функціональним блоком в межах мережевої інфраструктури, яка має чітко визначені зовнішні інтерфейси і чітко визначену функціональну поведінку. Прикладами мережевих функцій є елементи в домашній мережі, наприклад абонентський шлюз (Residential Gateway – RGW); і традиційні мережеві функції, наприклад, DHCP-сервери, брандмауери і т.д. Таким чином, VNF є реалізацією мережевої функції, яка розгорнута на віртуальних ресурсах, таких як віртуальна машина. Один VNF може складатися з декількох внутрішніх компонентів і, отже, він може бути розгорнутий на декількох віртуальних машинах, і в цьому випадку кожна віртуальна машина містить один з компонентів VNF. Сервіс є пропозицією від TSP, що складається з однієї або декількох мережевих функцій. У випадку NFV, мережеві функції, які утворюють сервіс, віртуалізуються і розгортаються на віртуальних ресурсах, таких як віртуальна машина. Проте, з точки зору користувачів, сервіси, що працюють на функціях спеціалізованого обладнання або на віртуальних машинах, повинні мати однакову продуктивність. Кількість, тип і порядок VNF, що створюють його,

визначаються функціональною і поведінковою специфікацією сервісу. Таким чином, поведінка сервісу залежить від складових VNF [12].

Мережевий сервіс архітектурно може розглядатися як граф мережевих функцій, з'єднаних між собою за допомогою відповідної мережевої інфраструктури. Ці мережеві функції можуть бути реалізовані в мережі одного оператора або у взаємодії між різними операторами мереж. Поведінка нижчерозташованої мережевої функції вносить вклад у поведінку сервісу вищого рівня. Таким чином, поведінка мережевого сервісу являє собою поєднання поведінки її складових функціональних блоків, які можуть включати в себе окремі мережеві функції, набори мережевих функцій, графи мережевих функцій і/або мережу інфраструктури.

Кінцеві точки і мережеві функції мережевого сервісу представлені у вигляді вузлів і відповідають пристроям, прикладним програмам і/або прикладним програмам фізичних серверів. Граф мережевих функцій може містити вузли мережевих функцій, з'єднаних логічними каналами, які можуть бути однонаправленими, двонаправленими, багатонаправленими та/або широкомовними. Простим прикладом графа є ланцюг мережевих функцій. Приклад такого мережевого сервісу з кінця в кінець може включати в себе смартфон, бездротову мережу, брандмауер, балансувальник навантаження і набір серверів мережі доставки контенту. Можливі віртуалізація і мережевий хостинг функцій абонента (наприклад, див. випадки використання Віртуальної Платформи як Сервіс (Virtual Network Platform as a Service – VNPaas) і віртуалізацію домашнього середовища в GS NFV 001 [22]).

Рис. 1.18 ілюструє представлення мережевого сервісу з кінця в кінець, що включає в себе другий вкладений граф мережевих функцій, як показано блоками вузлів мережевих функцій в середині рисунку, з'єднаних логічними каналами. Кінцеві точки підключаються до мережевих функцій через мережеву інфраструктуру (дротову або бездротову), в результаті чого ми бачимо логічний інтерфейс між кінцевою точкою і функцією мережі. Ці логічні інтерфейси представлені на рисунку пунктиром. На рис. 1.18,

зовнішній мережевий сервіс з кінця в кінець складається з кінцевої точки А, внутрішнього графа мережевих функцій і кінцевої точки В, в той час як внутрішній граф мережевих функцій складається з мережевих функцій NF1, NF2 і NF3. Вони з'єднані між собою за допомогою логічних каналів, що надаються мережею інфраструктури 2 [29].

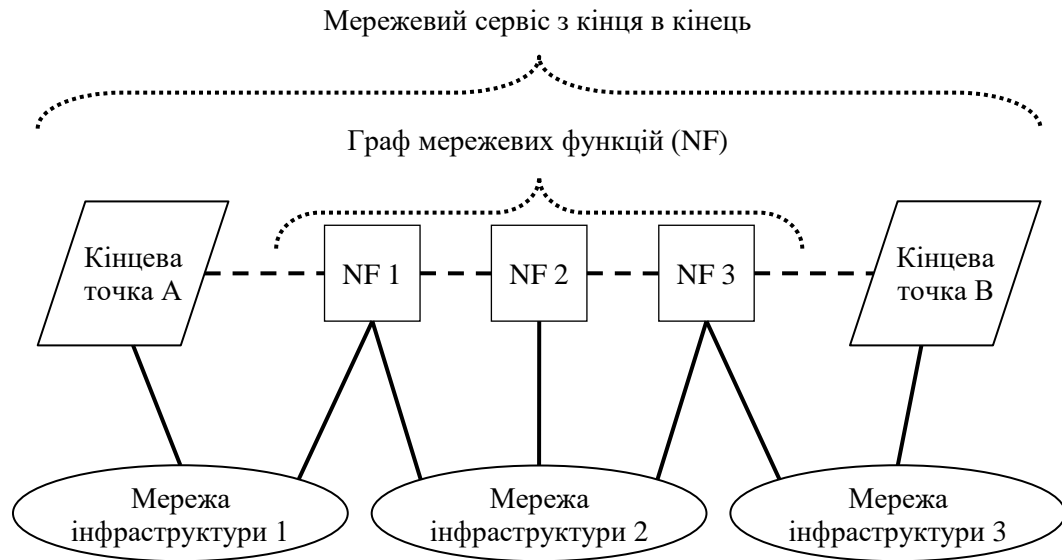


Рисунок 1.18 — Представлення мережевого сервісу з кінця в кінець у вигляді графа [29]

1.4.3 NFV Управління та Оркестровка (NFV Management and Orchestration – NFV MANO)

NFV MANO забезпечує функціональні можливості, необхідні для управління VNF, і пов'язаних з цим операцій, такі як конфігурація VNF та інфраструктури, на яких ці функції працюють. Вона включає в себе оркестровку і управління життєвим циклом фізичних і/або програмних ресурсів, що підтримують віртуалізацію інфраструктури, а також управління життєвим циклом VNF. Вона також включає в себе бази даних, які використовуються для зберігання моделей інформації і даних, які визначають як властивості розгортання, так і властивості життєвого циклу функцій, сервісів і ресурсів. NFV MANO фокусується на всіх завданнях управління віртуалізації необхідних платформі NFV. Крім того, платформа визначає

інтерфейси, які можуть використовуватися для зв'язку між різними компонентами NFV MANO, а також координації з традиційними системами керування мережею, такими як системи підтримки операцій (Operations Support System – OSS) і системи підтримки бізнесу (Business Support System – BSS), для того, щоб забезпечити керування як VNF, так і функціями, які працюють на традиційному обладнанні [12].

1.5 Хмарні обчислення

1.5.1 Загальна характеристика

Організація процесу надання послуг великій кількості користувачів, яка оцінюється в мільйони, має спільні риси. Так, все більше компаній користуються послугами хмарних сервісів, які забезпечені технічними ресурсами та можуть надавати інфраструктуру як сервіс [31].

Хмарні обчислення (Cloud Computing) – це модель забезпечення повсюдного та зручного доступу на вимогу через мережу до спільного пулу обчислювальних ресурсів, що підлягають налаштуванню (наприклад, до комунікаційних мереж, серверів, засобів збереження даних, прикладних програм та сервісів), і які можуть бути оперативно надані та звільнені з мінімальними управлінськими затратами та зверненнями до провайдера [32].

Для того, щоб розмістити віртуальну машину, фізична машина повинна надати всі ресурси, які вимагає віртуальна машина, в тому числі використання процесора, пам'ять, сховище та пропускну здатність мережевої карти [33].

Модель хмарних обчислень складається з п'яти ключових характеристик і трьох моделей обслуговування [34]. Коротко представимо їх далі.

1) Основні характеристики хмарних обчислень:

1. Самообслуговування на-вимогу. Споживач може самостійно керувати обчислювальними можливостями, такими як час сервера і мережеве сховище,

в міру необхідності автоматично без потреби у взаємодії людини з кожним сервіс провайдером.

2. Широкий доступ до мережі. Можливості (наприклад, обчислювальні ресурси, ємності зберігання) доступні через мережу і отримати до них доступ можна через стандартні механізми, які використовують різноманітні платформи тонкого або товстого клієнта (наприклад, мобільні телефони, планшети, ноутбуки і робочі станції).

3. Об'єднання ресурсів у пул. Обчислювальні ресурси провайдера об'єднують у пул для обслуговування багатьох споживачів з використанням мультитенантної моделі, з різними фізичними та віртуальними ресурсами, які динамічно призначаються і перепризначаються відповідно до споживчого попиту.

4. Швидка еластичність. Потужності можуть еластично надаватися і звільнятися, в деяких випадках автоматично, щоб швидко масштабуватися назовні і всередину пропорційно до вимог.

5. Вимірює обслуговування. Хмарні системи автоматично керують та оптимізують використання ресурсів за рахунок використання можливості вимірювання на певному рівні абстракції, відповідному типу обслуговування (наприклад, зберігання, обробки, пропускну здатності, і активних облікових записів користувачів).

2) Моделі обслуговування хмарних обчислень: три сервісні моделі хмарних обчислень показані на рис. 1.19, і визначені нижче [35]:

1. Програмне забезпечення як сервіс (Software as a Service – SaaS). Користувач має можливість використовувати прикладні програми провайдерів, що працюють на хмарній інфраструктурі. Прикладні програми є доступними з різних клієнтських пристроїв або через тонкий клієнтський інтерфейс, такий як веб-браузер (наприклад, веб-пошта) або інтерфейс програмування.

2. Платформа як сервіс (Platform as a Service – PaaS). Користувач має можливість розгорнути у інфраструктурі хмари створені споживачем або

придбані прикладні програми, створені з використанням мов програмування, бібліотек, сервісів і інструментів, підтримуваних провайдером.

3. Інфраструктура як сервіс (Infrastructure as a Service – IaaS). Користувач має можливість управляти ресурсами обробки, зберігання, мереж та іншими фундаментальними обчислювальними ресурсами, де споживач має можливість розгорнути і запустити довільне програмне забезпечення, яке може включати в себе операційні системи і прикладні програми.

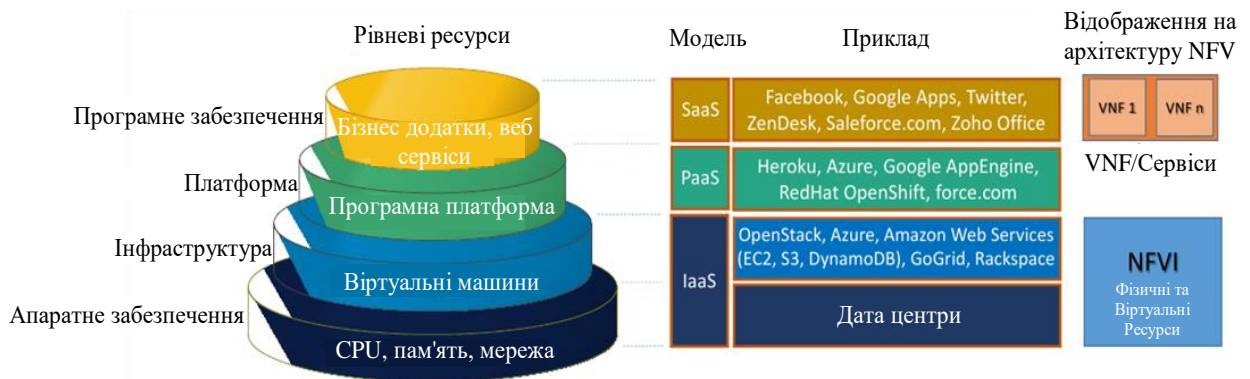


Рисунок 1.19 — Сервісні моделі хмарних обчислень та їх відображення на частину еталонної архітектури NFV

1.5.2 Взаємозв'язок між хмарними обчисленнями та NFV

Розгортання мережевих функцій в хмарі, ймовірно змінить всі аспекти, пов'язані з тим як сервіси і прикладні програми розробляються і поставляються. У той час як продовжуються робити відносно мережевих хмар і взаємодії між хмарами [35], [36], телекомунікаційні мережі відрізняються від середовища хмарних обчислень принаймні трьома способами: (1) навантаження площини даних в телекомунікаційних мережах мають високий тиск на продуктивність, (2) топології телекомунікаційних мереж накладають жорсткі вимоги на мережу і потребу в глобальному погляді на мережу для керування [37], (3) телекомунікаційна галузь вимагає масштабованості, доступності п'ять дев'яток та надійності. У традиційних телекомунікаційних мережах, ці характеристики забезпечуються інфраструктурою сайту. Якщо NFV засновується на хмарних обчисленнях, ці

функції повинні бути відтворені за допомогою хмарної інфраструктури таким чином, що вони можуть бути оркестровані, а оркестрові функції можуть бути доступні через відповідні абстракції, а також бути поєднані з розширеною підтримкою для виявлення і простежуваності [38]. У зв'язку з цим слід підкреслити, що NFV поставе більше задач, ніж просто перенесення функцій мережі операторського класу в хмару. Існує необхідність адаптувати хмарні середовища, щоб отримати поведінку операторського класу [37]. У таблиці 1.1 підсумовується зв'язок між NFV для телекомунікаційних мереж і хмарними обчисленнями [12].

Таблиця 1.1 — Порівняння NFV у телекомунікаційних мережах та хмарних обчисленнях

Проблема	NFV (телекомунікаційні мережі)	Хмарні обчислення
підхід	абстракція сервісу/функцій	абстракція обчислень
формалізація	ETSI NFV група стандартів промисловості	робоча група управління хмарами DMTF [39]
затримка	очікування низької затримки	деяка затримка допустима
інфраструктура	гетерогенний транспорт (оптичний, безпроводний, Ethernet)	гомогенний транспорт (Ethernet)
протокол	багато протоколів управління (наприклад OpenFlow [40], SNMP [41])	OpenFlow
надійність	суворі вимоги доступності 5 Дев'яток [42]	менш суворі вимоги надійності[43]
регуляція	суворі вимоги, наприклад NEBS [43]	різноманітні та змінювані

1.5.3 NFV, SDN та хмарні обчислення

Щоб підвести підсумок взаємозв'язку між NFV, SDN і хмарними обчисленнями, використовуємо рис. 1.20 (варто зауважити, що OpenFlow не єдиний протокол SDN. Точно так, OpenStack не єдина платформа хмарних обчислень. Причиною наведення лише цих двох на рис. 1.20 є те, що вони отримали більше уваги в цілому, а також і по відношенню до NFV) [12]. Зауважимо, що кожне з цих полів є абстракцією різних ресурсів: обчислювальних для хмарних обчислень, мережевих для SDN, і функцій для

NFV. Переваги, які отримуються від кожного з них схожі: гнучкість, зниження витрат, динамічність, автоматизація, масштабування ресурсів і т.д.

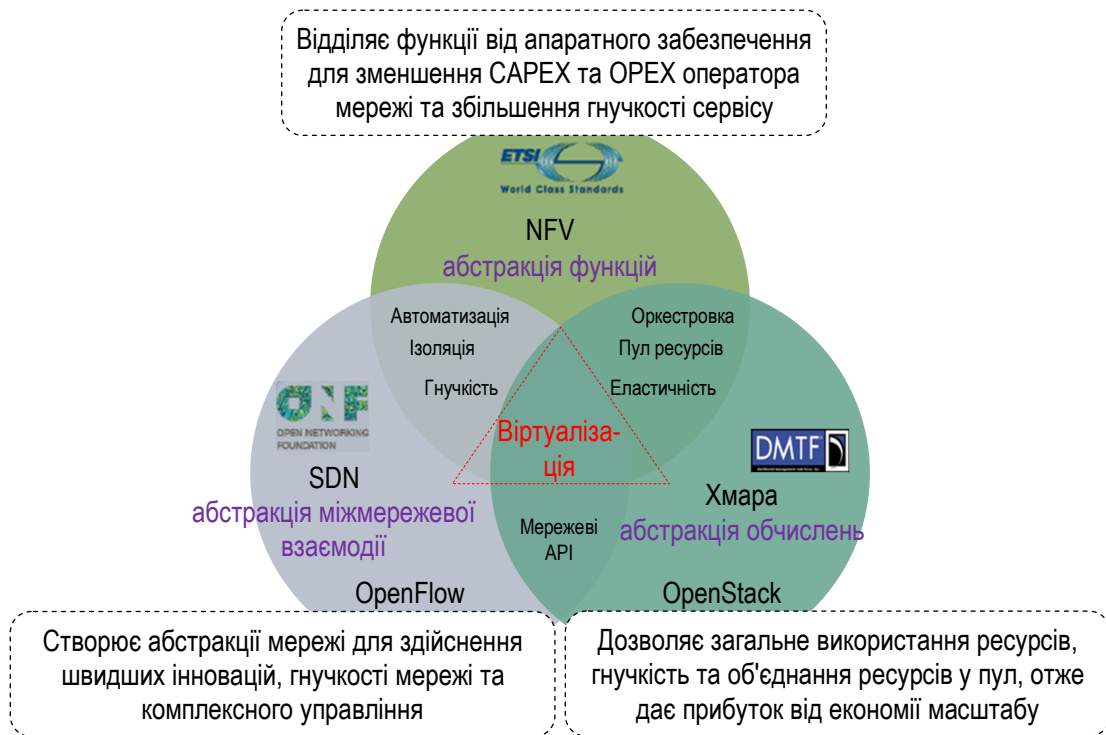


Рисунок 1.20 — Взаємозв'язок між NFV, SDN та хмарними обчисленнями

Питання не в тому, чи будуть мережеві функції перенесені в хмару, так як це насправді загальна ідея NFV. Воно в тому чи це буде хмара публічна, як Amazon, або якщо телекомунікаційні сервіс провайдери віддадуть перевагу використанню приватних хмар, розподілених по всій їх інфраструктурі. У будь-якому випадку, необхідно створити хмару операторського класу з точки зору продуктивності, надійності, безпеки, зв'язку між функціями і т.д.

З іншого боку, цілі NFV можуть бути досягнуті за допомогою механізмів, що не відносяться до SDN, і спираються на методи, що використовуються в даний час в багатьох дата центрах. Проте, підходи, що спираються на поділ площин управління і передачі даних як пропонується SDN, можуть підвищити продуктивність, спростити сумісність з існуючими реалізаціями, а також спростити процедури експлуатації та технічного обслуговування. Таким же чином, NFV здатний підтримувати SDN, надаючи

інфраструктуру, на якій програмне забезпечення SDN може бути запущене. І, нарешті, сучасний варіант дата центру (хмари і їх аспект самообслуговування) спирається на автоматизовані системи керування, які можуть бути отримані з SDN та NFV. Зокрема, такі аспекти, як мережа як сервіс, балансування навантаження, брандмауер, VPN і т.д. всі працюють в програмному забезпеченні що запускається за допомогою API.

1.6 Опис мережевого сервісу

NFV дозволяє операторам масштабувати послуги мережі з більшою деталізацією та оперативністю, ніж сьогодні, коли продуктивність мережевого сервісу визначена статично на найбільший прогнозований пік, що веде до надлишковості ресурсів. Для цього автоматизація є ключовою. У пошуках цієї автоматизації інститут ETSI визначив еталонну архітектуру NFV, яка використовує шаблони, керовані моделями, що називаються NSD (Network Service Descriptor) для забезпечення роботи мережевих сервісів. Для операції масштабування, NSD визначається дискретний набір рівнів функціонування, серед яких екземпляр мережевого сервісу може бути змінений у розмірі упродовж життєвого циклу [45].

VNF Descriptor (VNFD) – це шаблон специфікації, що надається VNF Провайдером для опису вимог віртуальних ресурсів VNF. Він використовується функціями NFV MANO для визначення способу виконання операцій життєвого циклу VNF (наприклад, створення екземпляра) [46].

На рис. 1.21 [47] показано спрощений приклад дескриптора VNF (VNFD) та його зв'язок з VNF. Приклад на рис. 1.21 показує екземпляр VNF, який складається з 4 компонентів VNFC (Virtualised Network Function Component) 3 різних типів: 'A', 'B' та 'C'. Кожен тип VNFC має власні вимоги до операційної системи (ОС) та середовища виконання (наприклад, віртуальної машини). Ці віртуальні ресурси та їхні вимоги до зв'язку описуються в елементах моделі даних, які формують VNFD. Окрім вимог до

ресурсів, VNFD також містить однозначні посилання на файли двійкового коду VNF, скрипти, дані конфігурації тощо, які необхідні для управління NFV та функцій оркестровки для належного налаштування VNF.

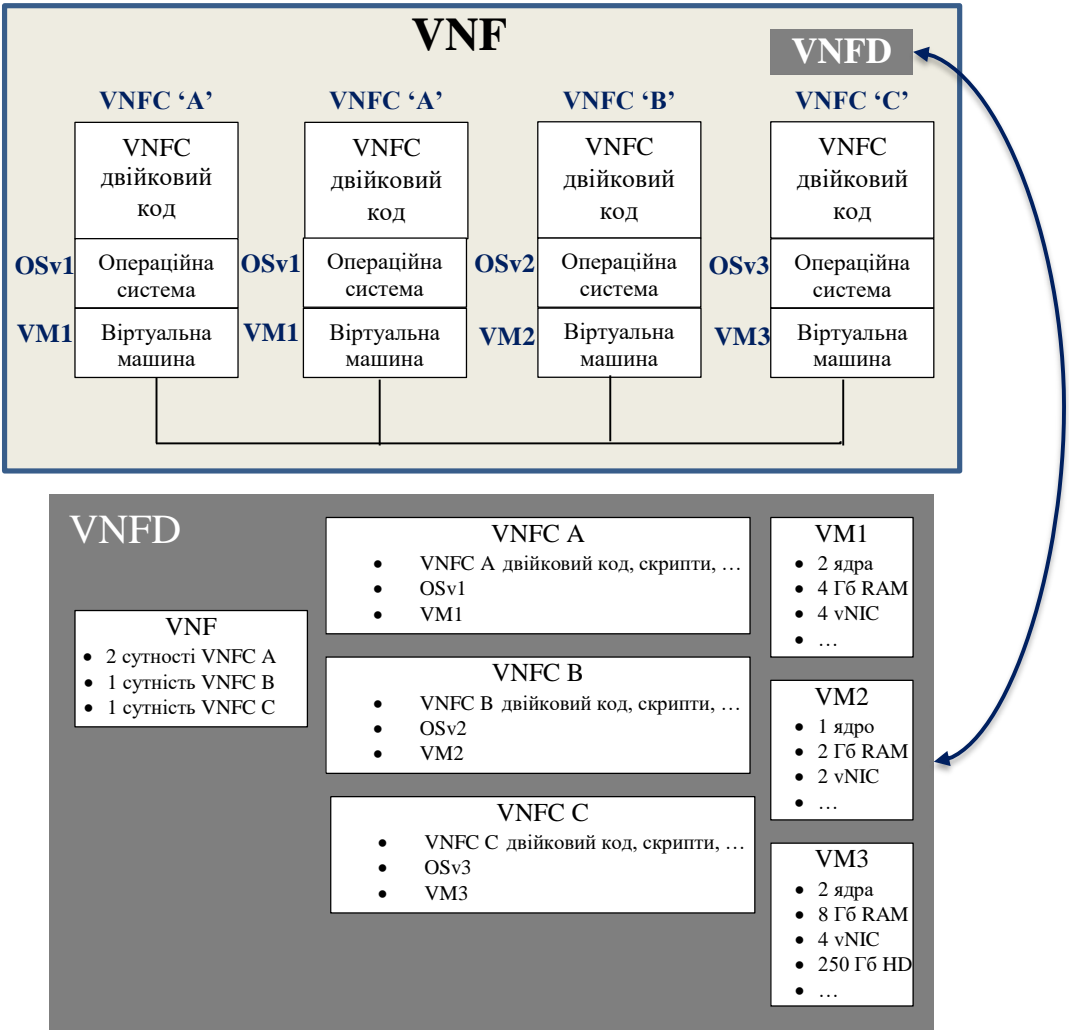


Рисунок 1.21 — Приклад елементів VNFD [47]

Вимоги до ресурсів NFVI (наприклад, вимоги до з'єднання, пропускної спроможності, затримки тощо) не показані на рис. 1.21, але передбачається, що вони присутні в VNFD, а також в інших дескрипторах, які використовуються функціями керування та оркестрування NFV. VNFD може також вказувати правила розташування (наприклад, деякі екземпляри VNFC повинні розміщатися на віртуальних машинах, які надаються ресурсами, розташованими в одній стійці).

Функції керування та оркестрування NFV враховують всі доступні атрибути VNFD, щоб перевірити можливість створення екземпляру даного VNF, наприклад. перевіряють, які типи ресурсів потрібні для кожного екземпляра VNFC. Є кілька варіантів того, як можна створювати екземпляри окремих VNFC, а саме:

- повністю або частково завантажені контейнери віртуалізації;
- порожні контейнери для віртуалізації, підготовлені для завантаження.

Наприклад, якщо використовується останній варіант, то файл завантаження контейнера віртуалізації буде NULL. Потім відповідальністю функцій керування та оркестрування VNF (наприклад, NFVO та VNF Manager) буде наказати VIM створити порожній контейнер віртуалізації з відповідним інтерфейсом SWA-5 (Vn-Nf), який готовий до використання.

Таким чином, Дескриптор Мережевого Сервісу (NSD) – це шаблон розгортання, який містить машинно оброблювану інформацію, що використовується блоками MANO, для створення екземплярів мережевого сервісу (NS) та керування ними протягом усього життя. NSD побудований з набору атрибутів та інших дескрипторів (рис. 1.22) [45].

Інформація моніторингу: визначає інформацію, яку слід відслідковувати для керування продуктивністю та несправностями NS. Ця інформація включає показники ефективності, пов'язані з ресурсами (на рівні NS/VNF) та індикатори VNF від складових VNF.

Правила автоматичного масштабування: містять правила, які дозволяють запускати дії по масштабуванню на екземплярі NS, коли умова, що містить інформацію моніторинга, не виконується. Інформаційна модель NFV дозволяє виражати ці правила як індивідуальні скрипти, що надаються під час створення екземпляру.

Особливості розгортання: описують конкретні конфігурації розгортання для NS.

Для опису розгортання та операційної поведінки компонентів NS, NSD містить і звертається до набору дескрипторів, включаючи дескриптори VNF, VL-дескриптори (VLDs) та дескриптори VNFFG (VNFFGDs) [48]. VNFD містить інформацію, необхідну для розгортання та функціонування екземплярів VNF. VLD надає інформацію про VL, включаючи конфігурації розгортання, доступні для створення екземпляру VL. Ці конфігурації визначаються за допомогою особливостей розгортання. Різні конфігурації забезпечують різні рівні продуктивності та надійності VNF. Нарешті, VNFFGD посилається на VNFDs та VLDs для опису топології.

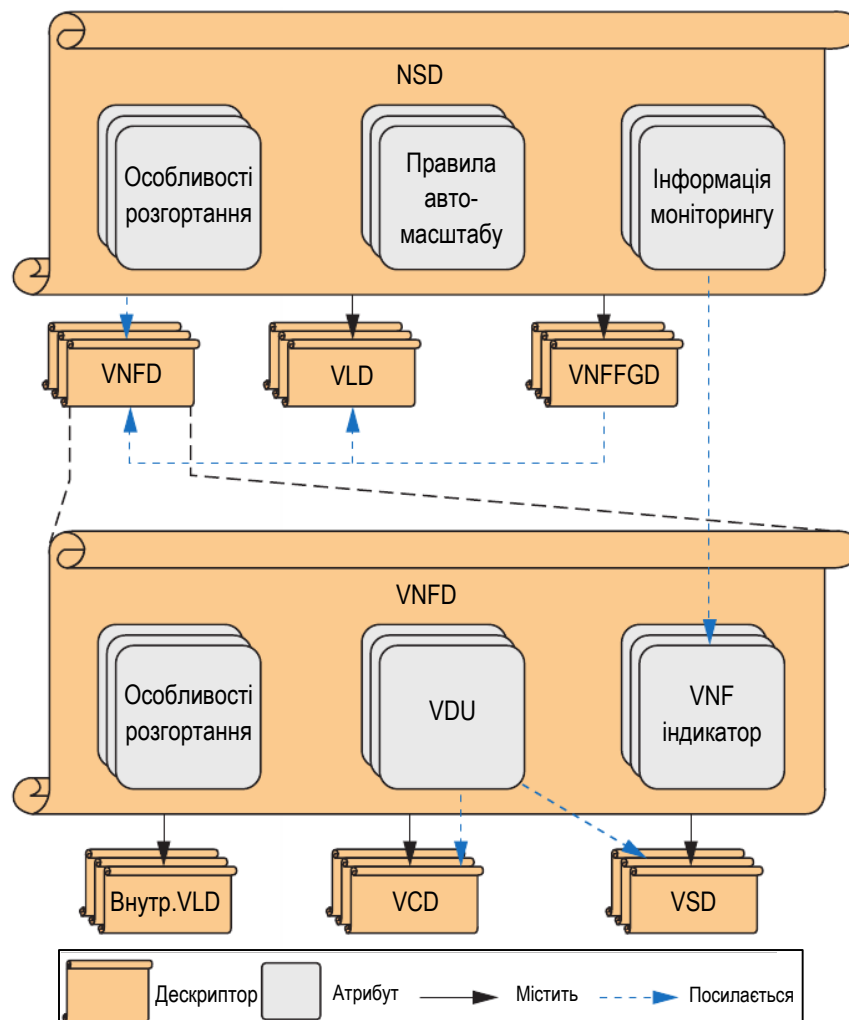


Рисунок 1.22 — Структура NSD. Показано дескриптори та атрибути найбільш релевантні для масштабування

З наведених дескрипторів розглянемо VNFD. NSD посилається на інформацію VNFD, яка є важливою для масштабування NS. Як і в NSD, VNFD також включає дескриптори та атрибути.

Дескриптори, що містяться в VNFD, містять детальний опис внутрішньої композиції VNF. Зокрема, VNFD включає в себе дескриптори віртуальних обчислень (virtual compute descriptors VCD), дескриптори віртуального зберігання (virtual storage descriptors VSD) та внутрішні VLD. Перші два вказують віртуальні ресурси обчислень та зберігання, необхідні для хостингу VNFC, тоді як останній визначає вимоги до продуктивності з'єднання VNFC (VNFC connectivity).

З точки зору атрибутів, VNFD включає в себе один або декілька наступних елементів:

Індикатори VNF: представляють події, пов'язані з продуктивністю/помилками, які надають інформацію про VNF на рівні прикладних програм.

Блоки розгортання віртуалізації (Virtualization Deployment Units VDU): описують як створювати та управляти екземплярами VNFC; отже, VDU можна розглядати як дескриптор VNFC. VDU визначає обчислювальні ресурси (і, можливо, ресурси зберігання), які необхідні контейнеру віртуалізації для розміщення VNFC. З цією метою він посилається на один VCD (і, можливо, один або більше VSD).

Особливості розгортання: подібні до тих, що визначені в NSD, але застосовні до VNFs.

Щоб автоматизувати запуск масштабування, NFVO має настроюваний програмний модуль (наприклад, що підтримує NS-специфічний код), який запускає алгоритм динамічного керування ресурсами (dynamic resource provisioning algorithm DRPA).

DRPA застосовує відповідні критерії оптимізації (наприклад, мінімізація витрат ресурсів, споживання енергії) та сукупність обмежень

(наприклад, доступна ємність ресурсів, обмеження місця розташування) до вихідних даних [45].

1.7 Особливості навантаження та роботи дата центра

Більшість існуючих досліджень проблеми вбудовування віртуальної мережі (Virtual Network Embedding – VNE) в літературі зосереджена на керуванні ресурсами шляхом резервування максимальні потреби в ресурсах для кожної віртуальної мережі протягом всього її життєвого циклу [49], [50].

Визначення ресурсів для сервісів у значній мірі залежить від точної оцінки характеристик обслуговування навантаження. Правильне керування ресурсами є складним завданням через коливання робочого навантаження.

Тим не менш, багато потоків заявок зазвичай мають періодичні шаблони (тобто, щодня, щотижня і/або сезонні цикли). Рис. 1.23 представляє графічно часові ряди запитів протягом п'яти днів, зібраних з програми для підприємства SAP, багаторівневої прикладної програми для бізнесу під назвою VDR і веб-прикладної програми, відповідно [51].

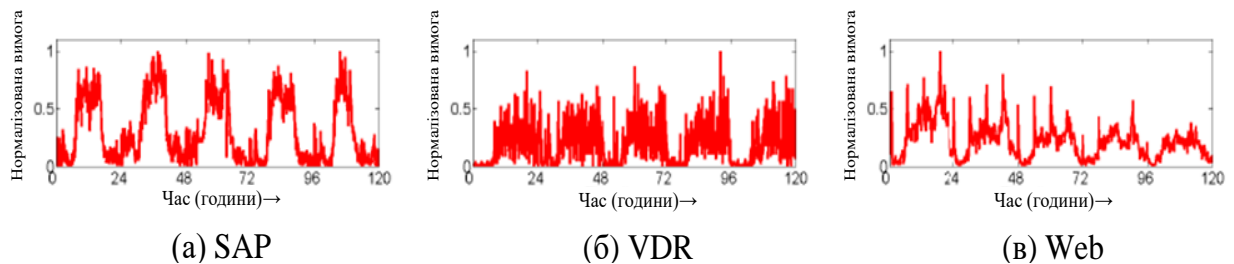


Рисунок 1.23 Часові ряди для трафіку (а) SAP; (б) VDR; і (в) Web [51]

Статичний метод керування ресурсами призводить до того, що до 85 відсотків хмарних ресурсів недовикористовуються більшість часу, що призводить до величезних втрат ресурсів і економічних втрат [52]. Завдяки використанню властивості періодичності, коефіцієнт використання ресурсів в хмарі може бути істотно поліпшено.

Проілюструємо це спостереження на прикладі, як показано на рис. 1.24, де Віртуальна мережа А надає офісним користувачам сервіси віртуальних настільних комп'ютерів, і яка як правило, відчуває низькі навантаження у вихідні дні, в той час як інша Віртуальна мережі В надає сервіси онлайн ігор і має високе навантаження в вихідні дні через високі вимоги користувачів. Якщо виділяти ресурси для А і В по їх максимальним вимогам, то тільки одна з них може бути відображена на фізичну мережу. Тим не менш, вони можуть бути обслужені, якщо їх змінні в часі потреби в ресурсах доповнюють один одного.

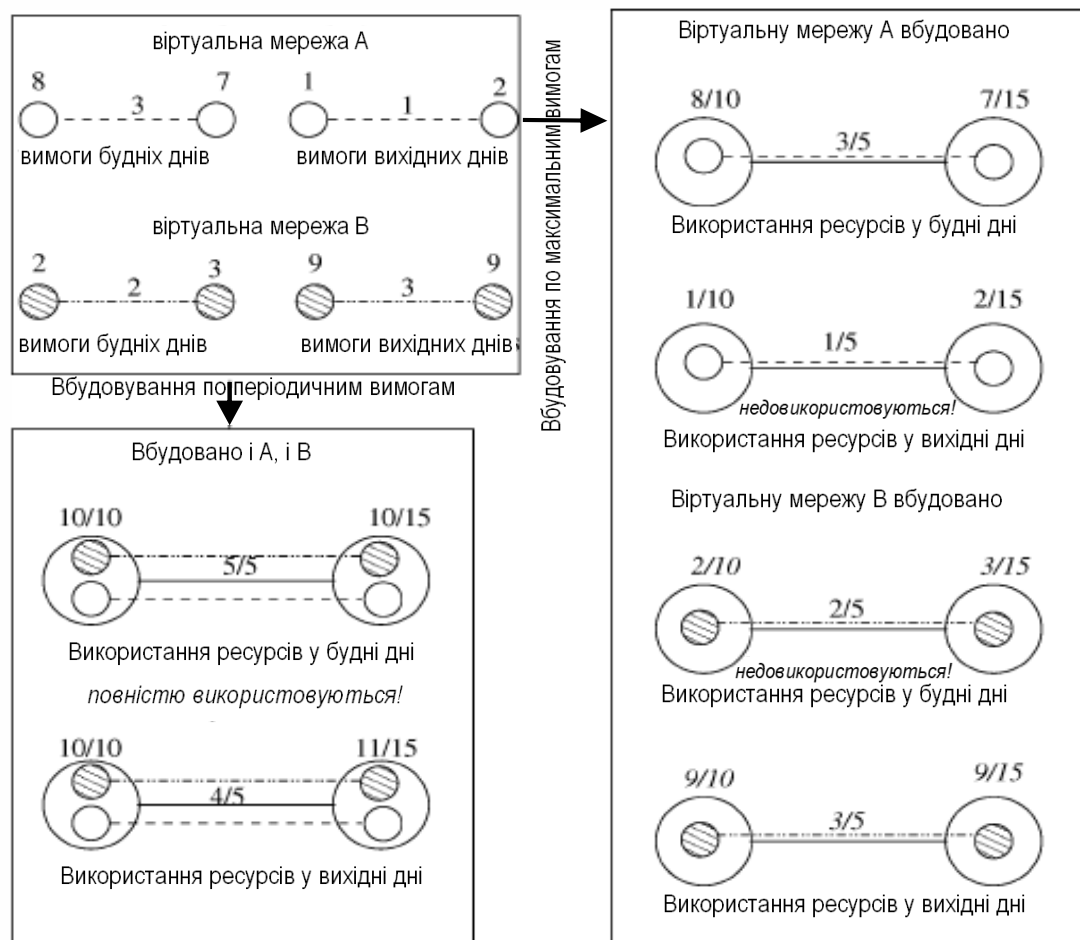


Рисунок 1.24 Приклад вбудовування мереж

Якщо визначити шаблони в навантаженні, тоді зможемо внести зміни відповідно до них в розподіл ресурсів і, отже, підвищити точність надання ресурсів і знизити енергоспоживання. Тим не менш, моделі вимог є

статистичними по своїй природі, і будуть мати місце відхилення від історичних моделей через непередбачені фактори, такі як несподівані натовпи, перебої в обслуговуванні та святкові дні. Хоча об'єм таких коливань малий в порівнянні з загальною вимогою, повне їх ігнорування може призвести до значних порушень угоди про якість послуг, що надаються, (Service Level Agreement – SLA) [51]. Через мінливість навантаження, що зазнають сучасні системи, розміщення віртуальних машин повинно постійно в режимі реального часу оптимізуватись [53].

Зрештою, надання ресурсів не є безкоштовним [51]; існують різні пов'язані з цим витрати і ризики. Часте виконання процедури надання ресурсів викликає як втрати продуктивності, так і енергії. Наприклад, ввімкнення серверу може зайняти значну кількість часу (до декількох хвилин) і споживати багато енергії (біля пікового споживання) [54]. Часті періодичні вмикання та вимикання потужності серверів викликає «знос», що може призвести до відмови сервера і перебоїв в обслуговуванні [55].

Є інші важливі проблеми, які виникають через високу потужність і енергоспоживання обчислювальних ресурсів. Потужність потрібна для забезпечення роботи охолоджуючої системи. На кожен Ват потужності, споживаний обчислювальними ресурсами, потрібно додатково 0.5-1 ват для системи охолодження [53]. Крім того, високе споживання енергії інфраструктурою призводить до суттєвих викидів діоксиду вуглецю (CO_2), що сприяють виникненню парникового ефекту.

Якщо потреби в ресурсах програми не задовільняються, програма може зіткнутися з підвищеним часом відповіді, тайм-аутами або відмовами. Забезпечення надійної якості обслуговування (QoS), визначеної за допомогою SLA має важливе значення для хмарних обчислювальних середовищ; тому, провайдери мають справу з компромісом по енергоефективності – мінімізацією споживання енергії з одночасним задоволенням SLA [53].

Оскільки навантаження на дата центри коливається з плином часу, можна було б вибірково вимикати частину системи для економії енергії, коли

вимоги на систему низькі. Енергозбереження досягається в результаті не тільки від переводу машин в стан сну, але і за рахунок економії витрат на охолодження [56].

Таким чином, навантаження, як правило, має значну мінливість. Це ускладнює надання ресурсів належним чином. Один обсяг (статичне виділення ресурсів) не може підійти всім випадкам, і призведе або до надмірного виділення ресурсів або до недостатнього виділення ресурсів [51].

1.8 Особливості дослідження

Хоча NFV обіцяє істотну економію коштів, гнучкість і простоту розгортання, потенційні проблеми в реалізації віртуалізованих мережевих елементів, які можуть підтримувати вимоги до продуктивності реального світу, і досі залишаються відкритим питанням [57], й в даний час NFV все ще перебуває на початкових етапах реалізації [58]. Предмет віртуалізованого ЕРС, що може динамічно реконфігуруватись, є досить новим сам по собі. Саме тому в цій сфері не так багато напрацювань [3].

Розподіл ресурсів в NFV мережах подібний до розміщення прикладних програм в дата центрах і хмарах [59]. Задача розміщення функцій також тісно пов'язана з вкладенням віртуальної мережі (VNE) [60]. В останні роки багато наукових досліджень [61] розглядали проблему VNE або застосовуючи математичні моделі оптимізації (наприклад, [2], [16], [62]) або алгоритмічні підходи (наприклад, [63]).

Наприклад, проблема розміщення контролерів розглядалась в [64] і [65], але з основним акцентом на досягнення мінімальної затримки SDN управління, а також стійкість надання послуг. У дослідженні [2] розглядається оптимальне розміщення дата центрів, на яких розміщуються віртуалізовані шлюзи, а також вирішується задача застосування віртуалізації та SDN декомпозиції на шлюзах мобільної базової мережі. Однак задача розміщення виникає і для інших функцій мережі, зокрема ЕРС. Подібно до

проблеми розміщення контролерів в області SDN, такі технології як NFV вимагають відповідних алгоритмів, які можуть вирішувати проблеми, що виходять за рамки однокритеріальних проблем розміщення. Ці проблеми можуть вносити додаткові складності через можливі взаємозалежності між мережевими функціями, як у випадку об'єднання функцій у ланцюжки, і нові потенційні обмеження відносно додаткових аспектів, таких як безпека [66].

Сучасні дослідження в основному розглядають тільки статичні запити віртуальної мережі, де запит і вимоги на ресурси є фіксованими і не змінюються з часом. Тим не менше, більшість запитів мають динамічні характеристики.

Далі ми розширюємо підхід VNE, визначаючи модель розташування мережеских функцій, що включає в себе поняття гібридних мереж, які містять як фізичні засоби обслуговування, так і віртуалізовані сервісні сутності. Схожим чином [16] також фокусується на розгортанні віртуальних мережеских функцій в гібридному середовищі, проте не враховує той факт, що продуктивність сервісної сутності залежить від виділених їй ресурсів. Також розглядається питання про те, як оптимально реконфігурувати розгорнуту віртуальну мережу в умовах зміни навантаження.

Також важливо відзначити, що наша робота відрізняється від більш загальних стратегій керування конфігурацією ресурсів в хмарних середовищах, оскільки ЕРС являє собою граф взаємозалежних вузлів, які не можуть розглядатися ізольовано [3].

Кілька дослідницьких праць були зосереджені на розробці адаптивних систем, які можуть реагувати на зміни у навантаженні в контексті систем зберігання, загальних операційних систем, мережеских сервісів, веб-серверів і Інтернет дата центрів [67]. У даному дослідженні ми розглядаємо абстрактну модель серверного ресурсу і представляємо методи динамічного розподілу ресурсів; запропоновані методи розподілу ресурсів застосовні до багатьох сценаріїв, де система або ресурс можуть абстрагуватись за допомогою GPS (Generalized Processor Sharing) сервера.

Одним з ключових аспектів у області віртуалізації мережі є виділення фізичних ресурсів віртуальним функціям мережі. Це передбачає відображення віртуальних мереж на фізичні мережі, а також керування виділеними ресурсами протягом усього життєвого циклу віртуальної мережі. Ефективність, оптимальність і гнучкість розподілу ресурсів є основними факторами для успішної віртуалізації мережі.

Вбудовування Віртуальної Мережі (VNE) є добре вивченою задачею [60]. Тим не менше, більшість сучасних рішень (наприклад, [68, 69]) пропонують статичну схему розподілу ресурсів, в якій коли віртуальна мережа відображається, перерозподіл ресурсів не відбувається протягом всього її життєвого циклу. Існує обмежена кількість децентралізованих і динамічних рішень VNE (як [70], що розглядає зміни в фізичній мережі, а не зміни в фактичному завантаженні віртуальних мереж, або [71], що пропонує алгоритми для задачі ефективної реконфігурації і вбудовування запитів віртуальної мережі, направлених у хмарний дата центр, автори вимагають, щоб провайдери направляли нові запити на зміну існуючих, і що тільки один такий запит може оброблятися в заданий момент часу). І навіть підходи, які пропонують рішення по динамічному вбудовуванню віртуальної мережі, все одно виділяють фіксовану кількість ресурсів для віртуальних вузлів і каналів на увесь період існування. Оскільки мережевий трафік не є статичним, це може привести до неналежного використання загальних мережевих ресурсів, особливо якщо фізична мережа відхиляє запити на вбудовування нових віртуальних мережевих функцій, при цьому резервуючи ресурси для віртуальних мережевих функцій, які знаходяться в умовах низької завантаженості [72].

Наше дослідження відрізняється від попередніх тим, що перерозподіл ресурсів є проактивним (не запускається невдалими вбудовуваннями), автономним (не запускається користувачами або мережевими провайдерами) і не передбачає повторні вбудовування вже відображених запитів.

Більшість існуючих робіт по динамічному керуванню ресурсами засновані на трьох підходах: теорії управління, моделюванні динаміки роботи і прогнозуванні навантаження [72]. Серед адаптивних систем, що використовують метод на основі теорії управління [67] і [73]. В системах, що застосовують даний підхід, як правило використовують попередньо визначену модель системи. На відміну від цього, наша методика заснована на онлайн характеристичній і прогнозуванні навантаження. Крім того, ці методи використовують залежність між параметром QoS (таким як цільова затримка) і параметром управління (таким як об'єм ресурсів), що не змінюється з часом. Серед робіт, що засновані на динаміці роботи – [74]. Автори [75] використовують прогнозування навантаження.

Підводячи підсумок, різниця між запропонованим у даному дослідженні підходом і згаданими вище є те, що в ньому ресурси, зарезервовані для використання віртуальними мережевими функціями не залишаються незмінними протягом усього життя віртуальної мережі. Здійснюється моніторинг віртуальних вузлів, і на основі їх реальних потреб у ресурсах, ресурси перерозподіляються, і в цьому випадку невикористані ресурси повертаються до фізичної мережі для використання іншими віртуальними мережами.

Наступною задачею, яка виникає, є те, яким чином отримується інформація про поточну ситуацію в мережі. В цьому аспекті керування конфігурацією ресурсів в NFV мережах подібне до керування прикладними програмами в дата центрах і хмарах.

Існуючі рішення керування ресурсами серверів можуть бути класифіковані як прогностичні і реактивні рішення. Прогностичне виділення ресурсів передбачає наявність передбачуваного і стабільного шаблону у вимогах і розподіляє об'єми, як правило, в масштабі часу декількох годин або днів на основі шаблону. У [76] використовується передбачуваність у вимогах бізнес прикладних програм, щоб підвищити ефективність керування ресурсами. Хоча ці підходи можуть бути ефективними в деякій мірі, обирати

правильний розмір інтервалу керування досі є дуже складним завданням. Запропонована нами методика дискретизації навантаження може допомогти з цим. Проте, великі, непередбачувані сплески вимог можуть викликати серйозні порушення SLA.

Реактивне виділення ресурсів, з іншого боку, виділяє ресурси в короткі проміжки часу (наприклад, кожні кілька хвилин) у відповідь на зміни навантаження (наприклад, [77]). Чисто реактивні політики потенційно можуть швидко реагувати на зміни навантаження, але такі проблеми, як непередбачуваність, нестабільність і високі витрати керування обмежують їх застосування на практиці [51].

Отже, прогнозувати пік навантаження прикладної програми та виділяти ресурси на основі оцінок найгіршого випадку вкрай складно [78]. З огляду на труднощі в прогнозуванні пікових навантажень, прикладна програма має використовувати комбінацію прогностичного і реактивного керування. У той час як прогностичні методи добре працюють для онлайн прогнозування на великих часових інтервалах до декількох годин, реактивні методи дозволяють прогнозувати навантаження на короткі часові інтервали до декількох хвилин і швидко реагувати на нестационарні перевантаження [67].

Існує кілька підходів, які поєднують в собі прогностичне і реактивне керування [51], [78]. Хоча ці підходи мають спільні риси з нашим підходом, вони розрізняються за кількома аспектами. По-перше, наш підхід спрямований на оптимізацію продуктивності і вартості виділення ресурсів одночасно. По-друге, пропонується метод аналізу та дискретизації навантаження, для визначення оптимального розміру інтервалів часу сталої конфігурації ресурсів – інтервали мають змінну довжину, тоді як в інших підходах до керування використовуються прості фіксовані інтервали [79]. Робота [51] схожим чином використовує змінну довжину інтервалів, однак на противагу цьому підходу, запропонований модифікований підхід довжину інтервалів визначає динамічно в залежності від фактичної ситуації в мережі.

Однією з основних проблем в динамічній мережі з NFV є те, як здійснити розумне розгортання VNF, щоб адаптуватися до змін в мережі. Маневрений механізм розгортання мережевих функцій може бути використаний для вирішення цієї проблеми. Коли змінюється мережа, адміністратор може видалити застарілі VNF і перерозподілити нові в кращому місці.

Існуючі роботи, пов'язані з розгортанням VNF і міграцією, як правило, зосереджені на пропозиціях нових стратегій розгортання [80] і механізмів міграції [81] [82]. Але вартість міграції не розглядається в цих дослідженнях. Насправді, вартість міграції є ключовим фактором в процесі міграції. Вона буде впливати на рішення про те, як вибрати кандидат VNF для перенесення і з'ясувати цільову позицію, яка підходить для міграції [83].

Стійкість обслуговування є важливою вимогою в будь-якій системі зв'язку, особливо в мобільних мережах, відомих своєю надійністю п'ять дев'яток. Крім того, доступність і надійність послуг, як зазначено в Рекомендації МСЕ-Т E.800 [84], визначають ключовий параметр характеристик якості обслуговування (QoS). В контексті хмарних мереж зв'язку, забезпечувати стійкість сервісів стає проблемою. Дійсно, висока доступність є важливою вимогою, але не обов'язково є невід'ємною рисою хмарних обчислень. Побудова системи, яка вимагає надійності п'яти дев'яток на платформі, яка не завжди може надати її, є таким чином перешкодою. В хмарному середовищі на стійкість обслуговування може в значній мірі впливати відмова будь-якої VNF, що працює на віртуальній машині. Відмова VNF може статися через кілька факторів, таких як відмова апаратних засобів (наприклад, через неправильне масштабування обладнання), вразливості програмного забезпечення та помилки в управлінні VNF (тобто, в основному, якщо VNF складається з декількох компонентів, кожен з яких розгорнутий на своїй власній віртуальній машині на тому ж або іншому апаратному забезпеченні) або її відповідній віртуальній машині, відмова на рівні гіпервізора через неправильну конфігурацію, негативний вплив на

продуктивність за рахунок інших VNF розміщених на одному фізичному вузлі, і зловмисних атак проти VNF або менеджера віртуальної машини (тобто, гіпервізора) [85]. В операторській хмарі, відмова VNF може вплинути на площину управління (наприклад, MME), а також на площину даних користувача (наприклад, S-GW або P-GW). У площині управління, роль MME має вирішальне значення, так як він відповідає за численні важливі процедури (наприклад, встановлення з'єднання з великим числом вузлів/VNF площини даних користувача, обладнання користувача (User Equipment – UE) – управління мобільністю і аутентифікацію UE). Його відмова істотно впливає на надання послуг, таким чином, важливість вивчення стійкості сервісів EPCaaS шляхом визначення оперативних, масштабованих і надійних механізмів відновлення для відновлення після збоїв VNF [86].

Одна з основних проблем в контексті віртуалізації мережі полягає в тому, як ефективно використовувати нижчерозташовані фізичні ресурси (тобто центральний процесор і пам'ять вузлів і пропускну здатність каналів). Відображення віртуальних вузлів і віртуальних каналів на фізичні ресурси, як відомо, є NP-важкою задачею [87].

Кілька дослідницьких спроб [88] вирішення цієї задачі були представлені; ряд цих досліджень вводив різні методи для вирішення завдання відображення віртуальної мережі в надії встановлення ефективного використання фізичних ресурсів. У доповненні до методів, необхідних для ефективного відображення віртуальних мереж на фізичні мережі, потрібні методи, які керують ресурсами, вже виділеними активним віртуальним мережам. Отже, існує необхідність розробити методику, яка може перемістити вже розміщені віртуальні вузли в разі відмови вузла або обслуговування вузла при зведенні до мінімуму періоду переривання обслуговування.

У той час як адаптація шляхів для віртуальних мереж була розглянута у ряді підходів [89], проблема відмови вузла в віртуальних мережах була розглянуто раніше лише у [90], проте не враховувалася вартість ресурсів на

вузлі і кінцева якість обслуговування, а також не розглянуто випадок відмови вузла через надмірне навантаження, що надходить на нього. Крім того, невирішеною залишалась задача вибору місць розташування вузлів керування.

У даній роботі, представляється розподілений алгоритм локальної реконфігурації мережі, який гнучко перерозподіляє віртуальні вузли, які постраждали від збою або перевантаження на фізичному вузлі. Основна мета полягає в розробці механізму самовідновлення віртуальної мережі, який може мінімізувати період переривання обслуговування і вартість відновлення вузла після відмови, а також підтримувати високий рівень фізичної працездатності мережі, що, в свою чергу, збільшує прибуток провайдера.

Таким чином, на відміну від існуючої статичної архітектури мережі LTE EPC пропонується система (рис. 1.25), в якій виділені апаратні мережеві функції обслуговують певний заданий рівень службового навантаження (штрихова лінія на рис. 1.26), в той час як заявки службових потоків, що цей рівень перевищують (крива на рис. 1.26), направляються на обробку з використанням віртуалізованих мережевих функцій у орендованих хмарах дата центрів.

Основні результати пошукових досліджень, викладені у розділі, наведено в роботах ([91-93]), де здобувачем систематизовано методи забезпечення якості функціонування складної багатокомпонентної телекомунікаційної системи та запропоновано моделі обслуговування сервісів в таких системах.

Далі описано задачу визначення місця розміщення та необхідного обсягу віртуальних зарезервованих ресурсів у контексті віртуалізації мережевих функцій. У розділі 3 описано підхід до керування інтервалом сталої конфігурації ресурсів з урахуванням прогнозу навантаження, в той час як пропонований метод для локальної реконфігурації ресурсів віртуалізованих мережевих функцій описаний у розділі 4.

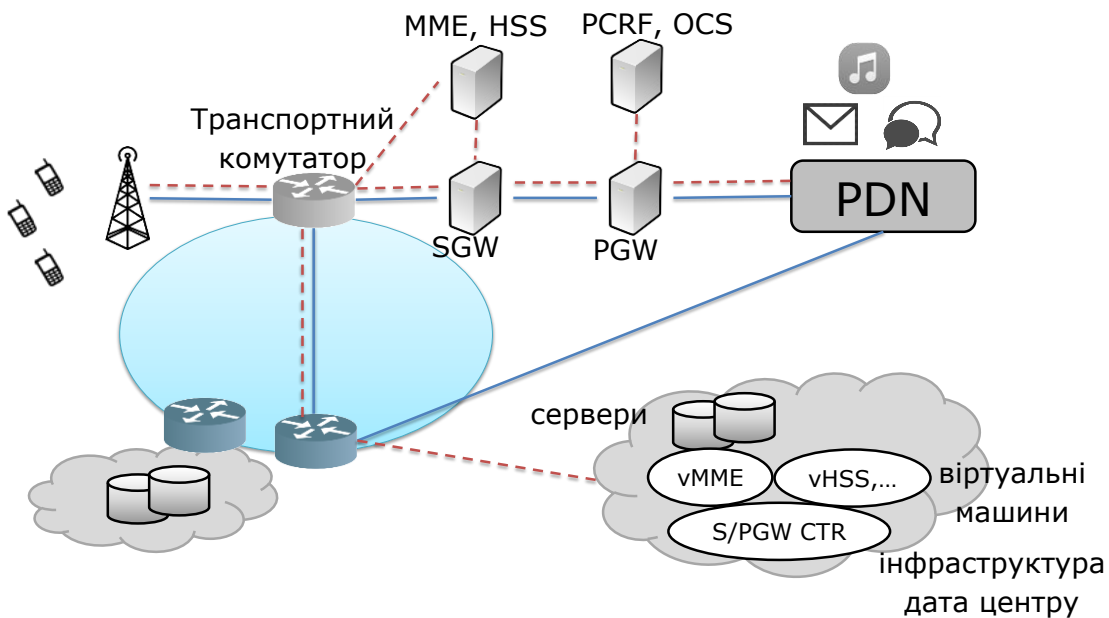


Рисунок 1.25 — Архітектура мережі LTE EPC з використанням NFV (змінна залежно від обсягів навантаження)



Рисунок 1.26 — Розподіл навантаження в гібридній мережі

Висновки

1. В результаті розгляду основних тенденцій технологій мобільних мереж майбутнього виявлено, що мобільні системи зв'язку наступних поколінь будуть застосовувати середовище з високими вимогами до якості зв'язку, тому для надання послуг і плавного переходу від одного покоління мобільних мереж до іншого, розроблена концепція реконфігурованості.

2. Визначення ключових характеристик застосування системи динамічного керування конфігурацією ресурсів показало доцільність її використання для вирішення актуальних телекомунікаційних проблем, хоча такі фактори, як підвищена складність роботи і проблеми регулювання в порівнянні з перевагами даного підходу повинні бути ретельно проаналізовані, перш ніж застосовувати його у системі мобільного зв'язку.

3. Проведений огляд сучасного телекомунікаційного середовища показав, що тенденції, такі як нестаріюче обладнання, безшовна інтеграція нових сервісів та багаторежимне обладнання у галузі комерційних безпроводових мереж привели до широкого інтересу до технології NFV.

4. Визначено, що для задоволення ряду суперечливих вимог до апаратних елементів необхідно застосовувати гібридні системи, які дозволяють спільно використовувати різні технології, що забезпечує найефективніше виконання кожної функції.

РОЗДІЛ 2

РЕЗЕРВУВАННЯ РЕСУРСІВ ДЛЯ ВІРТУАЛІЗОВАНИХ МЕРЕЖЕВИХ ФУНКЦІЙ В ГІБРИДНОМУ СЕРЕДОВИЩІ

2.1 Процедури LTE та потік заявок

Абонентське обладнання UE – це термінали, які дозволяють кожному користувачеві підключатися до мережі через базові станції eNodeB. На UE виконуються прикладні програми користувачів, які генерують або споживають мережевий трафік. UE може ініціювати запити до мережі, використовуючи керуючі повідомлення. Діяльність UE та генерація трафіку мережі також викликають процедури керування мережею. eNodeB отримують повідомлення UE до EPC.

Для забезпечення функціонування стандарт LTE визначає процедури сигналізації, які передбачають обмін сигнальними повідомленнями між об'єктами LTE (наприклад, eNB, MME, S-GW та HSS).

Для опису та розуміння того, як працює телекомунікаційна система, стандартизовано процедури мережі LTE. Детальний опис процедур викладено у технічних специфікаціях 3GPP TS 23.060 [94], 23.401 [95], і 23.402 [96]. Коротко розглянемо, для прикладу, процедуру підключення UE [97].

Підключення є першою процедурою, яку UE виконує після включення. Процедура виконується, щоб зробити можливим отримання послуг від мережі. Оптимізацією в системі LTE SAE (System Architecture Evolution) є те, що процедура підключення також включає в себе встановлення каналу EPS за замовчуванням, гарантуючи, що завжди на зв'язку IP з'єднання для UE/користувачів EPS доступне. Приклад процедури підключення описаний на рис. 2.1.

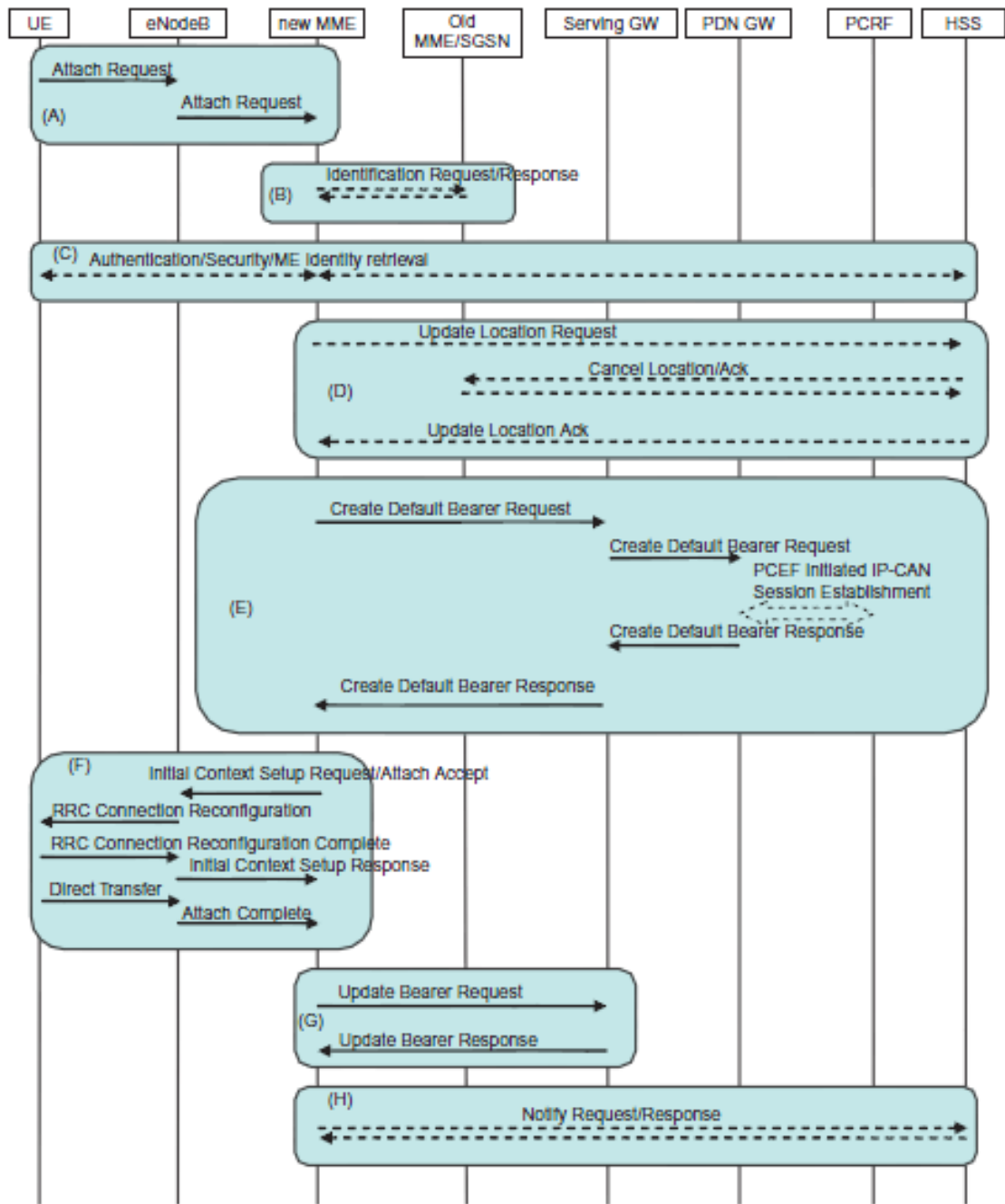


Рисунок 2.1 — Процедура підключення

Процедура коротко описана в наступних кроках:

A. UE відправляє повідомлення запиту Attach Request до eNodeB. eNodeB перевіряє ідентифікатор ММЕ переданий на рівні the Radio Resource Control (RRC). Якщо eNodeB має канал до визначеного ММЕ, він направляє запит Attach Request на цей ММЕ. Якщо не має, то eNodeB вибирає новий ММЕ і пересилає запит Attach Request.

В. ММЕ змінився і ММЕ використовує старий ММЕ ID в GUTI, щоб знайти старий ММЕ і отримує контекст UE.

С. Виконуються процедури аутентифікації та забезпечення безпеки. Визначення ME також виконується в поєднанні з цим кроком.

Д. Якщо ММЕ змінився, ММЕ повідомляє HSS, що UE перемістився. HSS зберігає адресу ММЕ і інструктує старий ММЕ скасувати контекст UE.

Е. Канал за замовчуванням авторизується PCRF і встановлюється між SGW і PGW.

Ф. Канал за замовчуванням встановлюється через радіоінтерфейс і підтвердження Attach Асерт надсилається в UE.

Г. ММА інформує SGW про ідентифікатор кінцевої точки тунелю (TEID) eNodeB, який завершує установку каналу за замовчуванням, так що він тепер може бути використаний як у висхідній лінії зв'язку, так і у низхідній лінії зв'язку.

Н. Якщо ММЕ вибрав PGW, який не є таким же, як той, що вказано в отриманій інформації підписки, то він надішле повідомлення про новий PGW на HSS.

Крім того, існують деякі додаткові кроки, які можуть бути виконані разом з процедурою підключення. Наприклад, якщо тимчасовий ідентифікатор UE (GUTI) невідомо як в старому ММЕ, так і в новому ММЕ (після кроків А і В), новий ММЕ проситиме UE відправити свій постійний ідентифікатор підписки (IMSI), як показано на рис. 2.2.

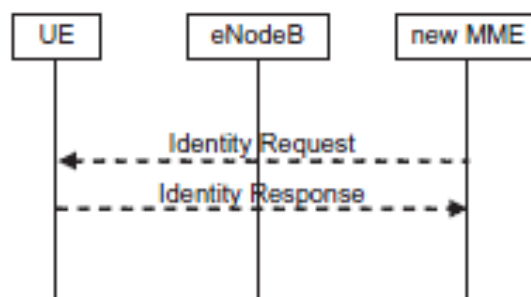


Рисунок 2.2 — Запит ідентифікатору

ММЕ може перевірити справжність МЕ за допомогою реєстру ідентифікації обладнання (EIR; після кроку С). EIR може бути використаний для чорного списку, наприклад, вкрадених UE. Залежно від відповіді від EIR, ММЕ може продовжити процедуру підключення або відхилити UE (див. рис. 2.3).

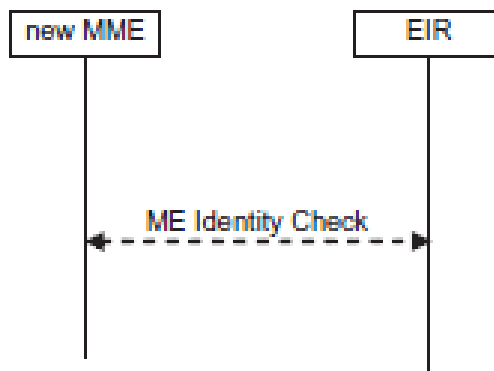


Рисунок 2.3 — Перевірка ідентифікатору

Якщо UE захотів відправити APN або PCO, він встановлює прапор в початковому повідомленні запиту Attach Request. ММЕ потім запитує інформацію від UE після початку шифрування на етапі С. Таким чином, немає ніякої необхідності відправляти APN або PCO у незашифрованому вигляді через інтерфейс радіозв'язку. Варіанти шифрування процедури запиту використовуються для передачі APN і/або PCO в ММЕ (див. рис. 2.4).

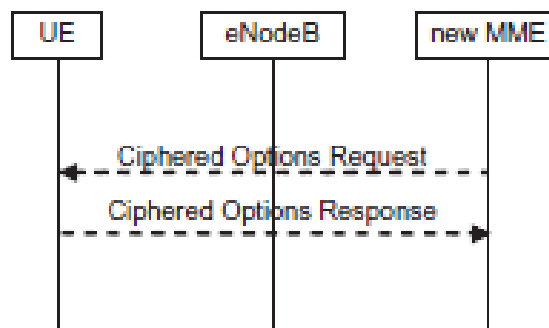


Рисунок 2.4 — Запит опцій шифрування

Потік заявок — це послідовність заявок, що надходять на обслуговування в систему в деякі моменти часу. Таким чином, кожен з типів

запитів процедур LTE формує пуасонівський потік заявок з деякими інтенсивностями $v_1, v_2, \dots, v_{proc}$. Маємо суму кінцевого числа $proc$ пуасонівських процесів, що формує сумарний пуасонівський потік з параметром інтенсивності $\lambda = v_1 + v_2 + \dots + v_{proc}$.

Середню інтенсивність вхідного потоку заявок λ до системи для первинного розміщення мережових функціональних блоків визначаємо виходячи з відомої середньої активності одного абонента (A):

$$\lambda = A \cdot Quantity,$$

де кількість абонентів $Quantity$ на території, що охоплюється, розраховується виходячи зі щільності населення ρ , що користується послугами обраного оператора: $Quantity = \rho \cdot S$, S – площа території.

Як приклад мережевої функції розглянемо блок MME. MME є основним елементом керування мережею. Він відповідає за збереження стану мобільності UE, управління каналом, аутентифікацію та авторизацію користувачів та інші функції. Коли MME одержує одне сигнальне повідомлення, він обробляє його, а пізніше надсилає нове повідомлення іншому об'єкту (такому як eNB або S-GW). Якщо процедура вимагає кількох етапів, сутність надсилає інше повідомлення до MME.

Існує кілька процедур сигналізації в LTE. З усіх них розглянемо ті, які генерують більшість сигнальних навантажень [98].

1) Service Request (SR): коли UE не має доступних ресурсів і створюється новий трафік, або від цього UE або від мережі до цього UE, UE виконує процедуру запиту сервісу Service Request (SR). Розглянемо SR, що запускається UE. Під час цієї процедури MME отримує три різні повідомлення (рис. 2.5): Initial UE Message (SR₁), Initial Context Setup Response (SR₂), Modify Bearer Response (SR₃).

Для обробки Initial UE Message (SR₁) MME повинен виконувати перевірку цілісності UE та дешифрування повідомлення. Крім того, він генерує ідентифікатори для каналів, які мають бути встановлені. Крім того, він зберігає та надає параметри та змінні, пов'язані з контекстом UE. Деякі з

них включаються в наступне повідомлення Initial Context Setup Request. Під час обробки повідомлення Initial Context Setup Response (SR_2), MME також отримує інформацію про контекст UE та включає цю інформацію в наступне повідомлення Modify Bearer Request. Обробка Modify Bearer Response (SR_3) є мінімальною, оскільки це повідомлення є лише підтвердженням.

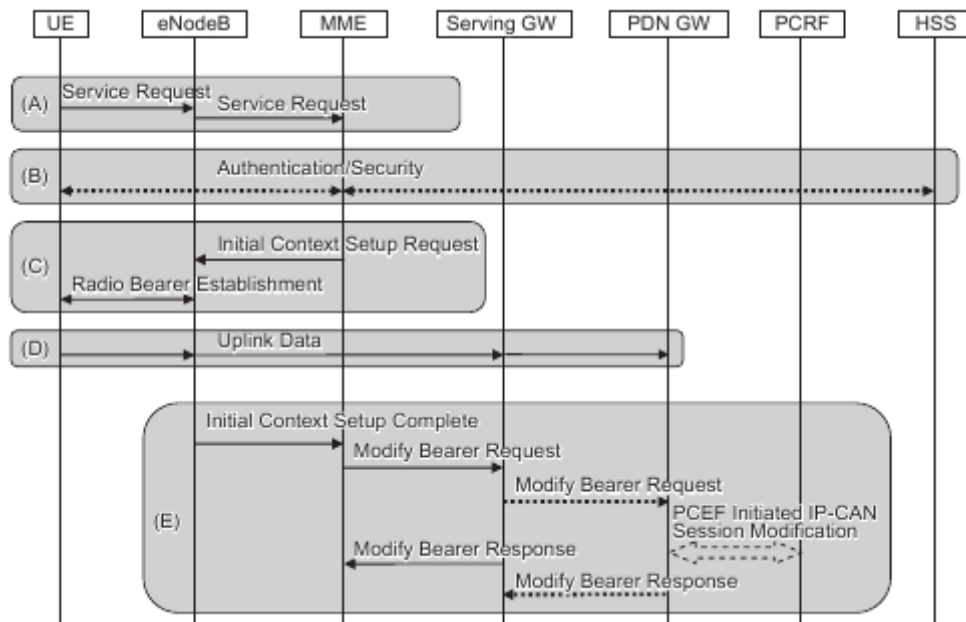


Рисунок 2.5 — Процедура «Запит сервісу» ініційована UE

2) Service Release (SRR): процедура Service Release (SRR) запускається неактивністю користувача. Її мета полягає в тому, щоб звільнити радіоканал даних та низхідний S1 канал у площині даних, а також сигнальні з'єднання радіо та S1 у площині управління для UE. Під час SRR MME обробляє три повідомлення (рис. 2.6): UE Context Release Request (SRR_1), Release Access Bearers Response (SRR_2), UE Context Release Complete (SRR_3).

Для обробки повідомлень UE Context Release Request message (SRR_1) та Release Access Bearers Request (SRR_2), MME має отримати інформацію про контекст UE та включити цю інформацію в наступні повідомлення. Обробка повідомлення UE Context Release Complete (SRR_3) в основному означає видалення контекстної інформації каналу.

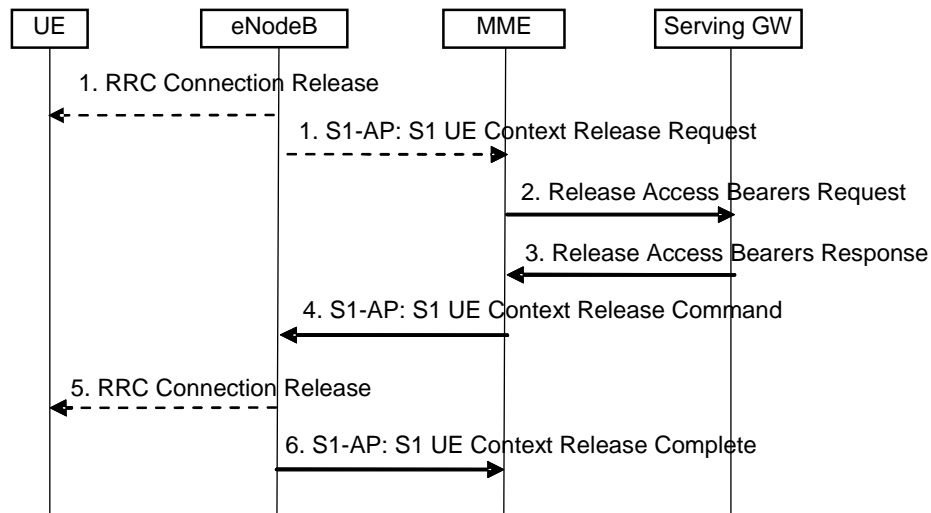


Рисунок 2.6 — Процедура «S1 звільнення»

3) X2-Based Handover (HR): MME бере участь у X2-Based Handover під час фази завершення хендоверу. Його мета – переключити кінцеву точку каналу від вихідної до цільової eNB. У цьому етапі MME отримує два повідомлення: Path Switch Request (HR₁) та Modify Bearer Response (HR₂) (рис. 2.7).

Для обробки цих повідомлень MME також повинен отримати інформацію про контекст UE і включити цю інформацію в наступні повідомлення. Для обробки повідомлення Path Switch Request, MME також необхідно зберігати нову інформацію, таку як ідентифікатори нового обслуговуючого стільника та нову область відстеження.

Нехай λ – сумарна інтенсивність надходження повідомлень сигналізації для блоку MME. Тоді, відповідно до опису процедур управління, λ розраховується як

$$\lambda = \nu_{SR} + \nu_{SRR} + \nu_{HR} = 3\lambda_{SR} + 3\lambda_{SRR} + 2\lambda_{HR},$$

де λ_{SR} , λ_{SRR} та λ_{HR} позначають середні значення інтенсивності надходження для процедур SR, SRR та HR, відповідно. Ці показники можна виразити в термінах середніх значень інтенсивностей для користувача Λ_U^h для процедури $h \in \{SR, SRR, HR\}$. Отримаємо $\lambda_h = \Lambda_U^h \cdot \text{Quantity}$.

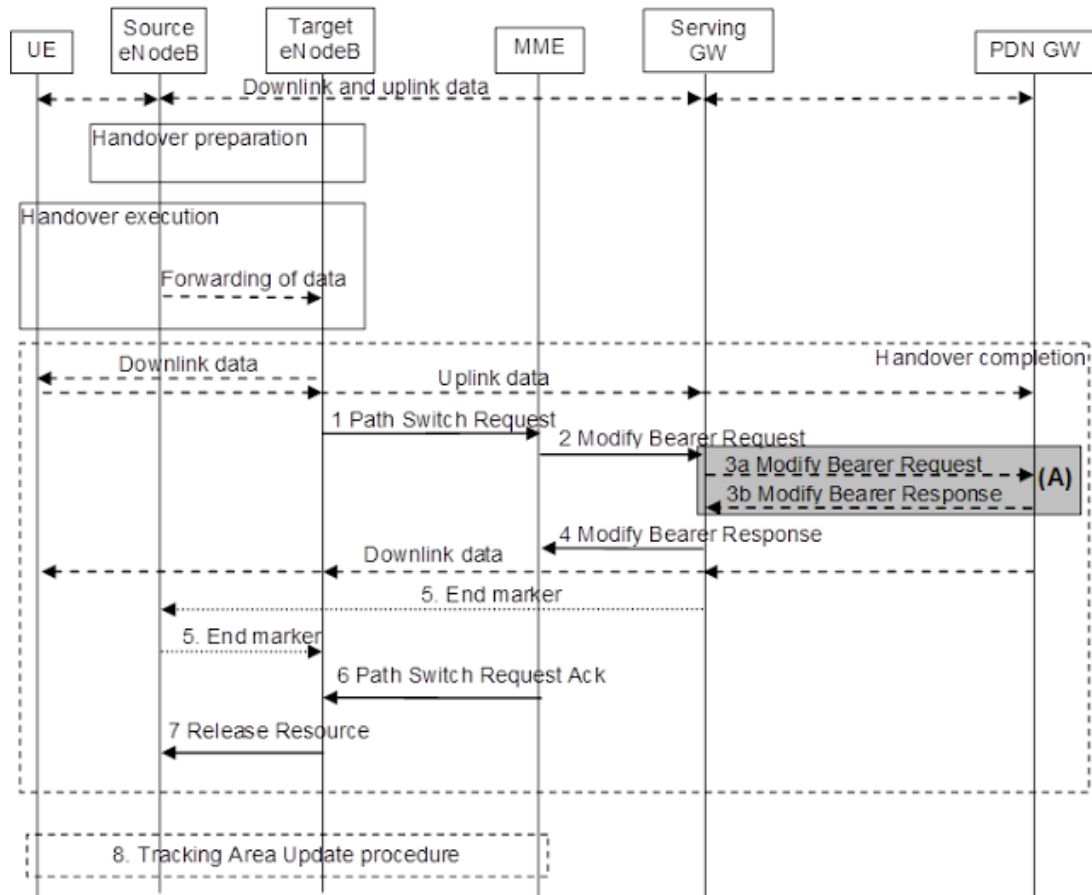


Рисунок 2.7 — Процедура «X2 хендовер без переміщення SGW»

2.2 Визначення кількості сервісів віртуальних мереж

Кількість сервісних ланцюгів потрібно визначити заздалегідь [61]. Крайнім випадком був б розгляд одного сервісного ланцюга на стільник/eNodeB. Оскільки реалістичні сценарії для мобільних мереж складають 10000 eNodeB, результуюча оптимізаційна модель буде величезна, і для її вирішення потрібен досить тривалий час обчислення. Тому приймаємо обґрунтовано великі кластери eNodeBs і припустимо, що кожен з цих кластерів eNodeB звертається до одного сервісного ланцюга базової мережі.

Більшість сучасних досліджень не розглядає задачу визначення кількості мереж EPS як сервісів, а передбачається їх наперед задана кількість. Найбільш близькою виступає задача з дослідження [99], яка, проте, направлена на визначення розміщення та прив'язки RRH сайтів до BBU

серверів у мережах з застосуваннях технології NFV і не враховує обмежень затримки та пропускної здатності.

Далі описано алгоритм вибору вузлів агрегації трафіку. Зокрема, розглядаємо випадок, коли провайдер телекомунікаційних послуг вже має існуючу топологію базових станцій. Потрібно визначити підмножину мережевих вузлів, де будуть розміщені блоки агрегації навантаження, які будуть формувати запити до одного віртуалізованого сервісу EPC. Після цього для кожного сайту базових станцій призначаємо вузол агрегації (Traffic Aggregation Point – TAP).

2.2.1 Постановка задачі вибору вузлів агрегації трафіку

Задача проілюстрована на рис. 2.8. На цьому рисунку певна кількість базових станцій групується та призначається до одного TAP.

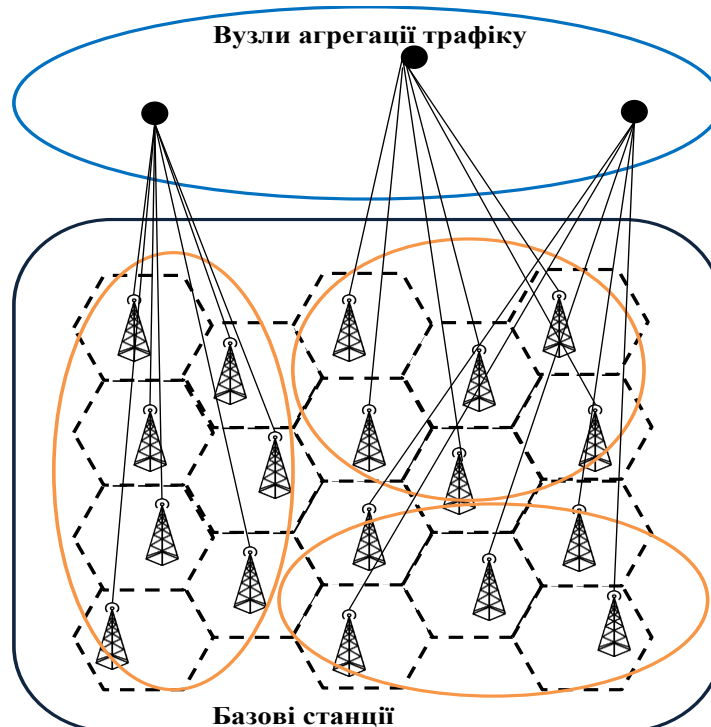


Рисунок 2.8 — Представлення проблеми розміщення вузлів агрегації

Сайти eNodeB

Розглядаємо мережу з множиною J сайтів eNodeB, які повинні обслуговуватись іншою множиною I вузлів агрегації, де $|I| \leq |J|$. Розташування

кожного сайту $j \in J$ характеризується географічною широтою та довготою. Кожен сайт базової станції $j \in J$ надає послуги абонентам, чії комбіновані середні вимоги до обробки є d_j . Для кожного сайту визначено максимальний рівень затримки T_j , який задає міру максимальної відстані від сайту до ТАР.

ТАР

Вважаємо, що провайдер може встановити максимум $p = |I| \leq |J|$ фізичних ТАР для обслуговування всіх сайтів мережі радіодоступу. Кожен ТАР $i \in I$ має максимальну ємність обробки P_i . Ця максимальна обчислювальна потужність спрямована на те, щоб відобразити той факт, що на практиці неможливо мати ТАР з необмеженим обсягом обчислювальних ресурсів. Розміщення ТАР $i \in I$ включає в себе вартість $cost_i$, яка може бути визначена як лінійна функція вартості, показана у (2.1). Вартість складається з двох частин: фіксованої початкової вартості f_i , яка відповідає за фіксовані інвестиції, такі як простір та встановлення обладнання, а також додаткові витрати $costN_i$ на одиницю обробної потужності, встановленої на обчислювальному вузлі, де dr_i – обсяг задіяних обчислювальних ресурсів обробки.

$$cost_i = f_i + costN_i \cdot dr_i \quad (2.1)$$

Канали зв'язку

Кожний сайт підключений до відповідного ТАР через лінію зв'язку лінії $e_{ji} \in E$, де E являє собою набір всіх ліній зв'язку. Кожна лінія зв'язку e_{ji} має затримку L_{ji} , яка залежить від відстані між сайтом j і ТАР i та швидкістю сигналів у використовуваному транспортному середовищі, а також доступну пропускну здатність c_{ji} . Вартість встановлення зв'язку між сайтом j і ТАР i дорівнює $costl_{ji}$ і визначається як лінійна комбінація початкової фіксованої вартості fl_{ji} та змінної частини залежної від пропускну здатності B_{ji} , необхідної каналу, та вартості одиниці пропускну здатності $costL_{ji}$, як показано в (2.2).

$$costl_{ji} = fl_{ji} + costL_{ji} \cdot B_{ji} \quad (2.2)$$

2.2.2 Формулювання оптимізаційної задачі

Визначення змінних рішення

Нехай x_i – це бінарна змінна рішення, яка приймає значення 1, якщо в точці i необхідно розмістити ТАР, та 0 – в іншому випадку. Крім того, визначаємо y_{ji} як бінарну змінну, яка приймає значення 1, якщо базова станція j направляє навантаження в i -у ТАР, і 0 – в іншому випадку. Необхідно визначити значення x_i та y_{ji} , так щоб знайти оптимальне значення цільової функції.

Цілі оптимізації

Цільова функція (2.3) прагне мінімізувати мережеві затримки.

Цільова функція (2.4) представляє загальну умовну вартість встановлення вузлів агрегації трафіку та вартість встановлення каналів між базовими станціями та відповідними ТАР, що їх обслуговують. При цьому:

$$cost_i = f_i + costN_i \cdot dr_i = f_i + costN_i \cdot \sum_j y_{ji} \cdot d_j.$$

Цільова функція (2.5) має на меті залишати більше вільної пропускної здатності на кожному фізичному каналі. Максимізується залишкова пропускна здатність по всіх каналах, оскільки значно завантажені канали можуть призвести до перевантажень у мережі, тому бажано отримати рішення, де залишається більше потужностей каналів.

Ці цілі оптимізації можуть бути корисними для мережевих операторів для планування найкращої стратегії розгортання.

$$\min_{x_i, y_{ji}} (\sum_i \sum_j y_{ji} \cdot L_{ji}), \quad (2.3)$$

$$\min_{x_i, y_{ji}} (\sum_i x_i \cdot cost_i + \sum_i \sum_j y_{ji} \cdot costl_{ji}), \quad (2.4)$$

$$\max_{x_i, y_{ji}} (\sum_i \sum_j y_{ji} \cdot (c_{ji} - B_{ji})) \quad (2.5)$$

Можливо використовувати лінійну комбінацію (2.6) виразів (2.3)-(2.5) з ваговими коефіцієнтами a , b , c , які можуть використовуватись не лише щоб задати більшу вагомість тої чи іншої складової, але також щоб масштабувати значення виразів з метою зведення до порівнюваних значень та мати значуще складення.

$$\min_{x_i, y_{ji}} \left(a \cdot \sum_i \sum_j y_{ji} \cdot L_{ji} + b \cdot \left(\sum_i x_i \cdot \text{cost} N_i + \sum_i \sum_j y_{ji} \cdot \text{cost} L_{ji} \right) - c \cdot \left(\sum_i \sum_j y_{ji} \cdot (c_{ji} - B_{ji}) \right) \right) \quad (2.6)$$

Обмеження

1. Розташування та призначення:

Обмеження (2.7) гарантує, що кожна базова станція буде приєднана до одного ТАР. Обмеження (2.8) гарантує, що між сайтом базової станції j та ТАР i створюється канал, тільки якщо i було розміщено. Разом обмеженнями (2.7) та (2.8) забезпечує, що розміщується потрібна кількість ТАР для обслуговування всіх сайтів, а базові станції призначаються лише вузлам, на яких розміщені ТАР.

$$\sum_i y_{ji} = 1 \quad \forall j, \quad (2.7)$$

$$y_{ji} \leq x_i \quad \forall j \quad \forall i \quad (2.8)$$

2. Забезпечення допустимості обсягу ресурсів:

Обмеження (2.9) гарантує, що максимальна кількість ТАР не перевищує бюджет p , тоді як (2.10) є обмеженням потужності, що гарантує, що загальні вимоги до обробки всіх базових станцій, призначених для конкретного ТАР, не перевищують фактичні встановлені фізичні ресурси. Обмеження (2.11) гарантує достатність каналних ресурсів для встановлення каналів, а (2.12) – допустимість значення затримки.

$$\sum_i x_i \leq p, \quad (2.9)$$

$$\sum_j y_{ji} \cdot d_j \leq p_i \quad \forall i, \quad (2.10)$$

$$\sum_i y_{ji} \cdot (c_{ji} - B_{ji}) \geq 0 \quad \forall j, \quad (2.11)$$

$$\sum_i y_{ji} \cdot L_{ji} \leq T_j \quad \forall j \quad (2.12)$$

Вирішити задачу (2.1)-(2.12) можна з застосуванням евристичних методів оптимізації (наприклад, генетичного алгоритму тощо).

Було здійснено оцінку роботи спрощеної моделі (2.6)-(2.12) порівняно з підходом направленим на мінімізацію затримки у середовищі моделювання Matlab. Приклад рішення задачі розміщення вузлів агрегації відповідно до запропонованого методу та методу найменшої затримки проілюстровано на рис. 2.9.

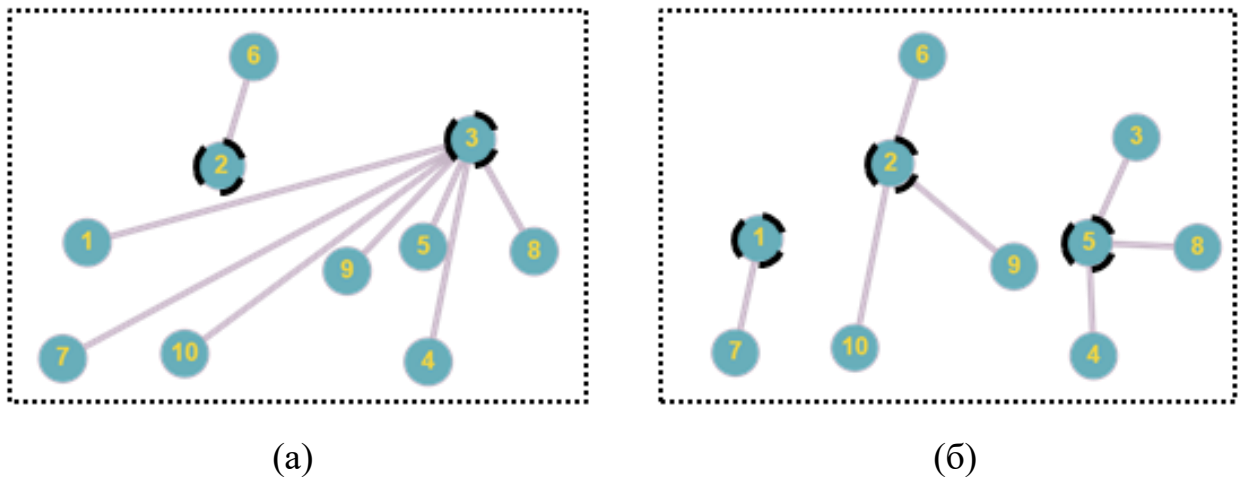


Рисунок 2.9 — Розміщення вузлів агрегації відповідно до запропонованого методу (а) та методу найменшої затримки (б)

Результати експериментів показали, що можливо скоротити витрати в два рази, проте ціною відповідного збільшення затримки.

Фрагмент програмного коду представлено на рис. 2.10.

```

k=1;
for j=1:10
    for i=1:10
        EqO=a*Lat(j,i)+e*costL(j,i)-c*(cbndw(j,i)-Bbndw(j,i));
        Obj(k,:)=EqO;
        k=k+1;
    end
end
for i=1:10
    EqO=e*costN(i);
    Obj(k,:)=EqO;
    k=k+1;
end

k=1;
for j=1:10
    for i=1:10
        AConstrEq(k,(j-1)*10+i)=1;
    end
    AConstrEq(k,110)=0;
    BConstrEq(k)=1;
    k=k+1;
end

k=1;
for i=1:10
    for j=1:10
        AConstrIneq(k,(j-1)*10+i)=1;
        AConstrIneq(k,100+i)=-1;
        BConstrIneq(k)=0;
        k=k+1;
    end
end

for i=1:10
    AConstrIneq(k,100+i)=1;
end
BConstrIneq(k)=p;
k=k+1;
for j=1:10
    for i=1:10
        AConstrIneq(k,(j-1)*10+i)=Bbndw(j,i)-cbndw(j,i);
    end
    BConstrIneq(k)=0;
    k=k+1;
end

for j=1:10
    for i=1:10
        AConstrIneq(k,(j-1)*10+i)=Lat(j,i);
    end
    BConstrIneq(k)=Tconstr(j);
    k=k+1;
end
end

```

Рисунок 2.10 — Фрагмент програмного коду задачі розміщення

2.3 Відображення віртуальних вузлів на фізичні вузли

Підхід базується на спільному розташуванні індивідуальних ланцюгів сервісів базової мережі на фізичній мережі, де ланцюг сервісів позначає мережеві функції мобільної базової мережі, яку потік трафіку повинен пройти [61]. Припускаємо, що віртуальні мережеві функції мають таку ж функціональність і інтерфейси як і мережеві елементи архітектури 3GPP LTE EPC.

Фізична мережа задана у вигляді графа $SN = (N, NE)$, де N є множиною фізичних вузлів і NE – множиною каналів. Кожен канал $l = (n_1, n_2) \in NE$, $n_1, n_2 \in N$ має максимальну пропускну здатність $c(n_1, n_2)$ і кожен вузол $n \in N$ пов'язаний з певними ресурсами c_n^i , $i \in R$, де R – множина типів ресурсів. Множина усіх точок агрегації трафіку (TAP), тобто кластерів eNodeB, в мережі позначається $K \subseteq N$. Для кожного вузла $n \in N$, $suit_n^{kj}$ є бінарним параметром, який вказує, чи адміністративно можливо розгорнути на вузлі функцію типу $j \in V$, де V є множиною типів мережевих функцій, k -го сервісу, де $k \in K$.

Віртуальна базова мобільна мережа представляється множиною сервісів (один сервіс на TAP), які вбудовуються в фізичну мережу.

Вимоги смуги пропускання каналу між двома функціями, $j1$ і $j2$, $(j1, j2) \in E$, що відносяться до сервісу TAP $k \in K$ позначається як $d_k^{(j1, j2)}$. $d_k^{j, i}$ – кількість ресурсу типу i , що виділяється для мережевої функції j сервісу k . $s_{n, i}^{kj}$ позначає час обробки запиту на ресурсі типу i мережевої функції j сервісу k однією одиницею ресурсу вузла n . Вимоги до допустимого часу обробки заявки мережевою функцією j , що відноситься до сервісу k , позначаються як P_k^j . T_k – максимальна затримка для $k \in K$, $L(n_1, n_2)$ – мережева затримка для каналу $(n_1, n_2) \in NE$.

Метою оптимізації є знаходження розташування віртуалізованих сервісів базової мережі (тобто розміщення мережевих функцій та розподіл

ресурсів, а також визначення шляхів передачі трафіку між ними), так щоб мінімізувати витрати на зайняті ресурси каналів і вузлів у фізичній мережі, при цьому задовільняючи вимоги трафіку. Сформулюємо цільову функцію (вираз (2.13)) у вигляді лінійної комбінації двох вартісних виразів: зайнятого обсягу ресурсів обчислювальних вузлів, де умовна вартість одиниці ресурсу i на вузлі n позначається як $costN(i,n)$, і зайнятої пропускної здатності каналів, де $costL(n_1,n_2)$ – умовна вартість одиниці пропускної здатності фізичного каналу $(n_1,n_2) \in NE$.

Наступні формули (2.13)-(2.22) представляють собою постановку оптимізаційної задачі нелінійного програмування. Змінні $x_n^{k,j}$ вказують, чи мережева функція j пов'язана сервісом k розташовується на фізичному вузлі n . Для $j=TAP$, $x_n^{k,TAP}$ – не змінні, а вхідні параметри, які вказують де TAP k знаходиться, тобто

$$x_n^{k,TAP} = \begin{cases} 1 & \text{якщо } k = n, \\ 0 & \text{інакше} \end{cases}.$$

Аналогічно, змінні $f_{(n_1,n_2)}^{k,(j_1,j_2)}$ вказують, чи фізичний канал $(n_1,n_2) \in NE$ використовується для шляху між j_1 і j_2 для сервісу k .

$$\min_{x_n^{k,j}, f_{(n_1,n_2)}^{k,(j_1,j_2)}, d_t^{j,i}} \left(\sum_{k \in K} \sum_{j \in V} \sum_{n \in N} \sum_{i \in R} x_n^{k,j} \cdot d_k^{j,i} \cdot costN(i,n) + \right. \\ \left. + \sum_{(n_1,n_2) \in NE} costL(n_1,n_2) \cdot \sum_{k \in K} \sum_{(j_1,j_2) \in E} f_{(n_1,n_2)}^{k,(j_1,j_2)} \cdot d_k^{(j_1,j_2)} \right) \quad (2.13)$$

$$\text{З обмеженнями } \sum_{n \in N} x_n^{k,j} = 1 \quad \forall k \in K, j \in V, \quad (2.14)$$

$$x_n^{k,j} \leq suit_n^{k,j} \quad \forall k \in K, j \in V, n \in N, \quad (2.15)$$

$$\sum_{(w,n) \in NE} \sum_{k \in K} \sum_{(j_1,j_2) \in E} f_{(w,n)}^{k,(j_1,j_2)} \cdot d_k^{(j_1,j_2)} \leq c_n^{bdw} \quad \forall n \in N, \quad (2.16)$$

$$\sum_{k \in K} \sum_{j \in V} x_n^{k,j} \cdot d_k^{j,i} \leq c_n^i \quad \forall n \in N, i \in \{R \setminus bdw\}, \quad (2.17)$$

$$\sum_{k \in K} \sum_{(j_1,j_2) \in E} f_{(n_1,n_2)}^{k,(j_1,j_2)} \cdot d_k^{(j_1,j_2)} \leq c(n_1,n_2) \quad \forall (n_1,n_2) \in NE, \quad (2.18)$$

$$\sum_{(n,w) \in NE} f_{(w,n)}^{k,(j_1,j_2)} - f_{(n,w)}^{k,(j_1,j_2)} = x_n^{k,j_1} - x_n^{k,j_2} \quad \forall k \in K, n \in N, (j_1,j_2) \in E, \quad (2.19)$$

$$x_n^{k,j}, f_{(n_1,n_2)}^{k,(j_1,j_2)} \in \{0,1\} \quad \forall k \in K, j \in V, n \in N, (j_1,j_2) \in E, (n_1,n_2) \in NE, \quad (2.20)$$

$$\sum_{(j_1,j_2) \in E} \sum_{(n_1,n_2) \in NE} f_{(n_1,n_2)}^{k,(j_1,j_2)} \cdot L(n_1,n_2) \leq T_k \quad \forall k \in K, \quad (2.21)$$

$$\sum_{n \in N} x_n^{k,j} \sum_{i \in R} \left(\frac{1}{\frac{d_k^{j,i}}{s_{n,i}^{k,j}} - \lambda^{k,j}} \right) \leq P_k^j \quad \forall t \in T, j \in V \quad (2.22)$$

Вираз (2.14) гарантує, що для кожної ТАР/сервісу розміщується тільки одна мережева функція кожного типу. Це забезпечується тим чином, що сума змінних $x_n^{k,j}$ для розташування мережевої функції j пов'язаної з сервісом k на фізичному вузлі n по всім вузлам дорівнює одиниці.

Вираз (2.15) гарантує, що розміщення ресурсів здійснюється на фізичних вузлах, які мають адміністративну можливість для розташування відповідних мережевих функцій; оскільки гарантується, що для випадку, коли булева змінна $x_n^{k,j} = 1$, змінна $suit_n^{k,j}$ буде відмінна від нуля.

Вирази (2.16), (2.17) і (2.18) являють собою обмеження на ресурси фізичних вузлів і каналів, тобто забезпечують той факт, що кількість задіяних на вузлі ресурсів не перевищує кількості наявних ресурсів. Слід зауважити, що канал між двома мережевими функціями відображається на шлях у фізичній мережі. Таким чином, його вимоги до пропускної спроможності впливають не тільки на мережеві ресурси фізичних вузлів, де мережеві функції розміщуються, але й на мережеві ресурси проміжних вузлів, які лежать на шляху (вираз (2.16)).

Вираз (2.19) представляє собою обмеження щодо збереження потоку для всіх шляхів у фізичній мережі, тобто що вхідний потік на вузлі дорівнює вихідному потоку.

Вираз (2.20) гарантує, що змінні у задачі розміщення функцій мережі та відображення шляху є булевими.

Щоб обмежити затримки на каналах, обмеження на затримку, показане в виразі (2.21), також додається. А для того, щоб урахувати у моделі необхідну продуктивність віртуальної мережевої функції, у виразі (2.22)

визначені обмеження на значення часу обробки заявки, що залежить від кількості виділених функціональному блоку ресурсів $d_k^{j,i}$ та часу обслуговування одиницею ресурсу $s_{n,i}^{k,j}$.

На рис. 2.11 представлено приклад системи відображення мережевих сервісів на фізичну мережу.

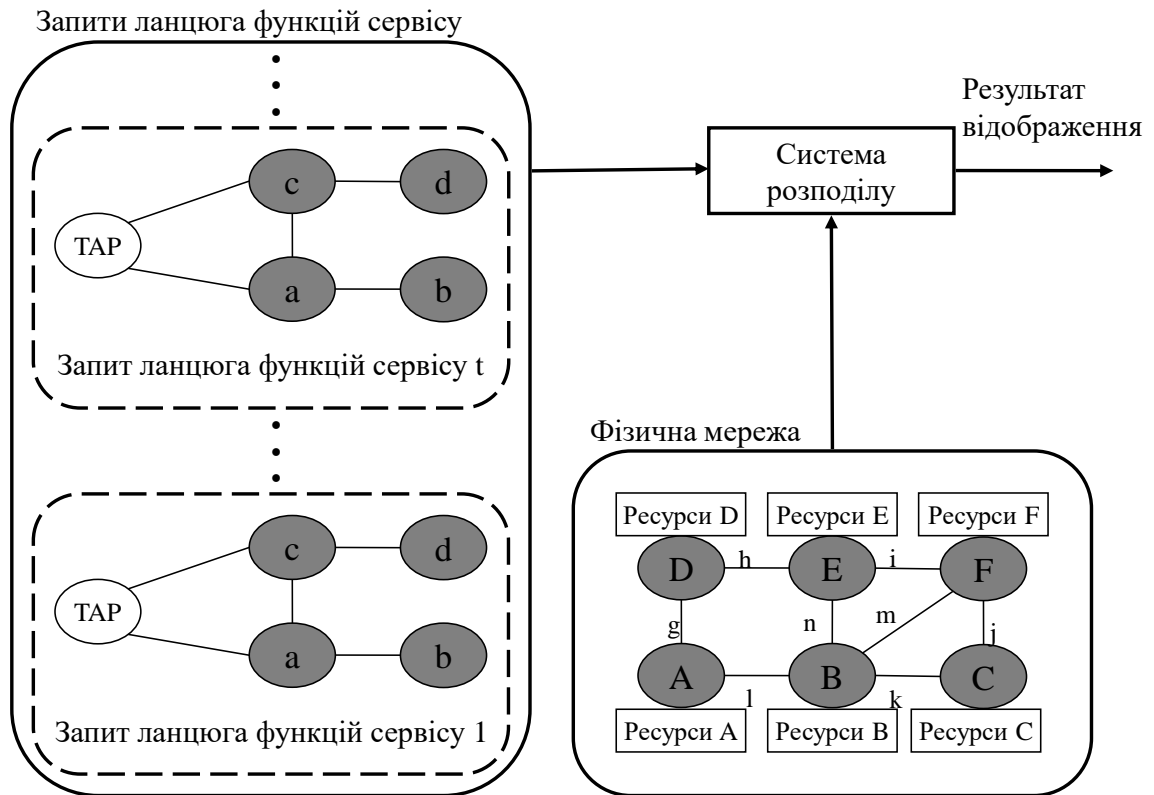


Рисунок 2.11 — Система виділення мережевих ресурсів – приклад топології

Передбачається вирішення задачі (2.13)-(2.22) в офлайн режимі на початковому етапі. Згідно з рішенням, кожній мережевій функції резервується певна кількість ресурсів віртуальної мережевої функції, на основі оцінки її найбільшої потреби в ресурсах; миттєві потреби різних мережевих функцій динамічно задовольняються шляхом активації необхідної конфігурації віртуальних машин під час виконання таким чином, щоб задовольнити гарантії передбачені для кожної мережевої функції.

Процес виділення ресурсів для віртуальних мережевих функцій зображено на рис. 2.12.

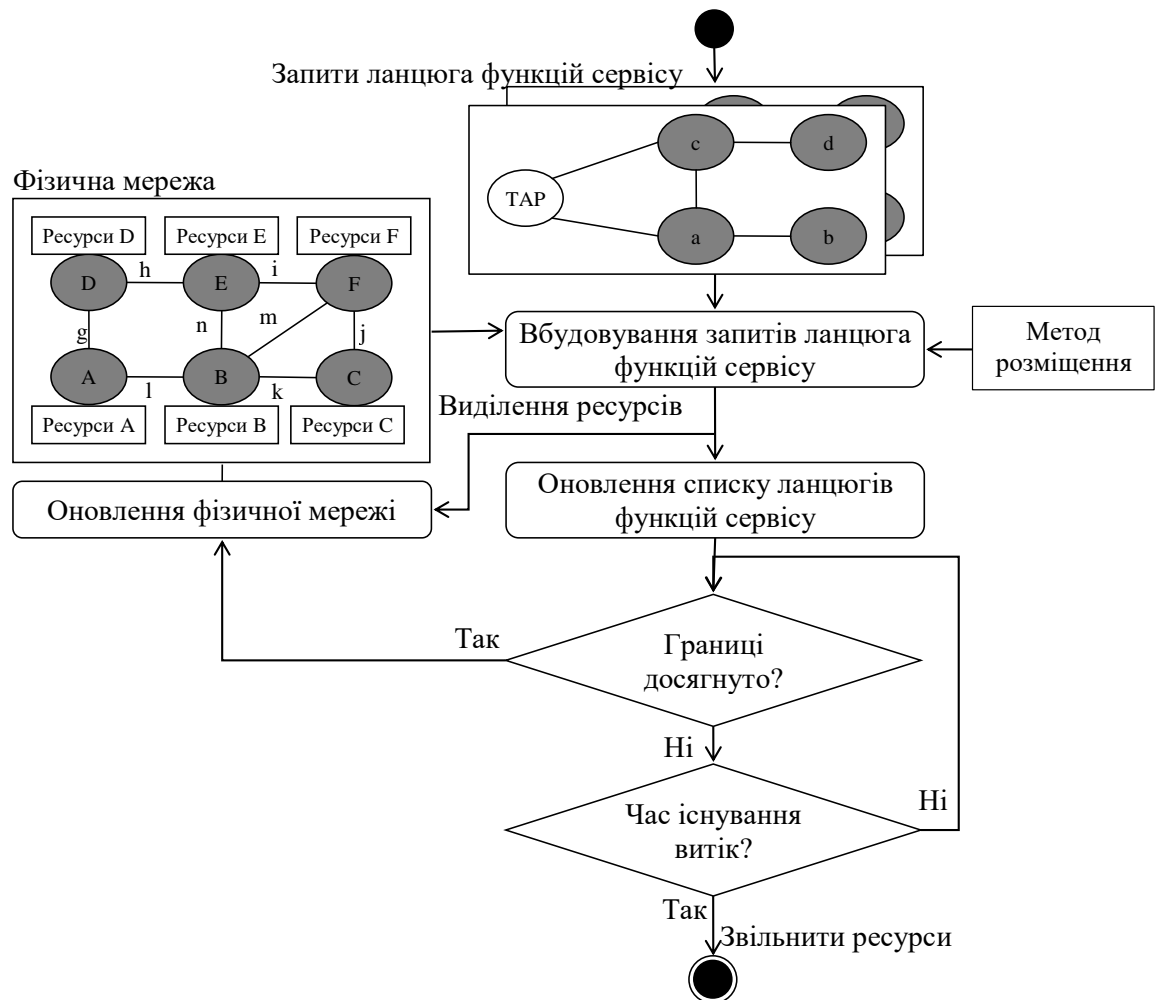


Рисунок 2.12 — Життєвий цикл запиту віртуальної мережі – діаграма активності

Процес виділення ресурсів починається з приходом запиту ланцюга функцій сервісу, як показано на рис. 2.12. Метод розміщення та резервування ресурсів початкового етапу (тобто (2.13)-(2.22)) використовується для вбудовування запитів ланцюга функцій мережі, він приймає на вхід поточний стан фізичної мережі (наприклад, доступні ресурси CPU, пам'яті та пропускну здатності) і самі запити.

Метод направлений на мінімізацію використання ресурсів, при цьому забезпечуючи заданий рівень якості обслуговування. Запропонований метод дозволяє телекомунікаційному оператору мінімізувати капітальні та

операційні витрати використовуючи парадигму хмарних обчислень та значно покращити якість сприйняття.

2.4 Еластичне керування EPC за допомогою поділу стану

Запропонована архітектура складається з межових вузлів (Edge Node – EN) і центрального вузла (Central Node – CN). CN виступає як традиційна мережа EPC, а також надає потенційно можливості хмарних обчислень, також підтримуючи глобальне уявлення про систему. Мережеві функції можуть розгортатися як в CN і так і в вузлах EN, користуючись усіма перевагами, які надходять від хмарних механізмів [100].

Основна ідея модифікованого методу полягає в ієрархічно розподіленій архітектурі (рис. 2.13), що дозволяє мінімізувати навантаження на мережу, зменшує навантаження на пристрої (процесор, пам'ять і енергію) і прискорює процес функціонування. Для цього вводиться проксі менеджер синхронізації (Proxy Synchronization Manager – PSM). PSM виступає в якості проксі сутності для переговорів з іншими сутностями мережі, зокрема, обслуговуючим менеджером синхронізації (Serving Synchronization Manager – SSM) і домашнім менеджером синхронізації (Home Synchronization Manager – HSM). SSM масштабується відповідно до очікуваних сервісів мережі і містить інформацію стану стосовно сусідніх EN. SSM передає інформацію управління до PSM. HSM є центральним вузлом, який зберігає інформацію щодо стану всієї мережі. HSM і SSM є ключовими елементами для управління великою кількістю вузлів і вони мають вирішальний вплив на продуктивність мережі.

Розглянемо тепер необхідність покращеної координації для забезпечення передачі стану між EN. Не представляється можливим використовувати існуючі процедури EPC і повинні бути введені деякі додаткові механізми. Серед кількох можливих альтернативних рішень, ми вибрали для поділу станом проактивну передачу стану і синхронізацію зміни

стану у сусідніх EN випереджувальним чином, всякий раз, коли нова стан з'являється в системі або оновлюється (рис. 2.14).

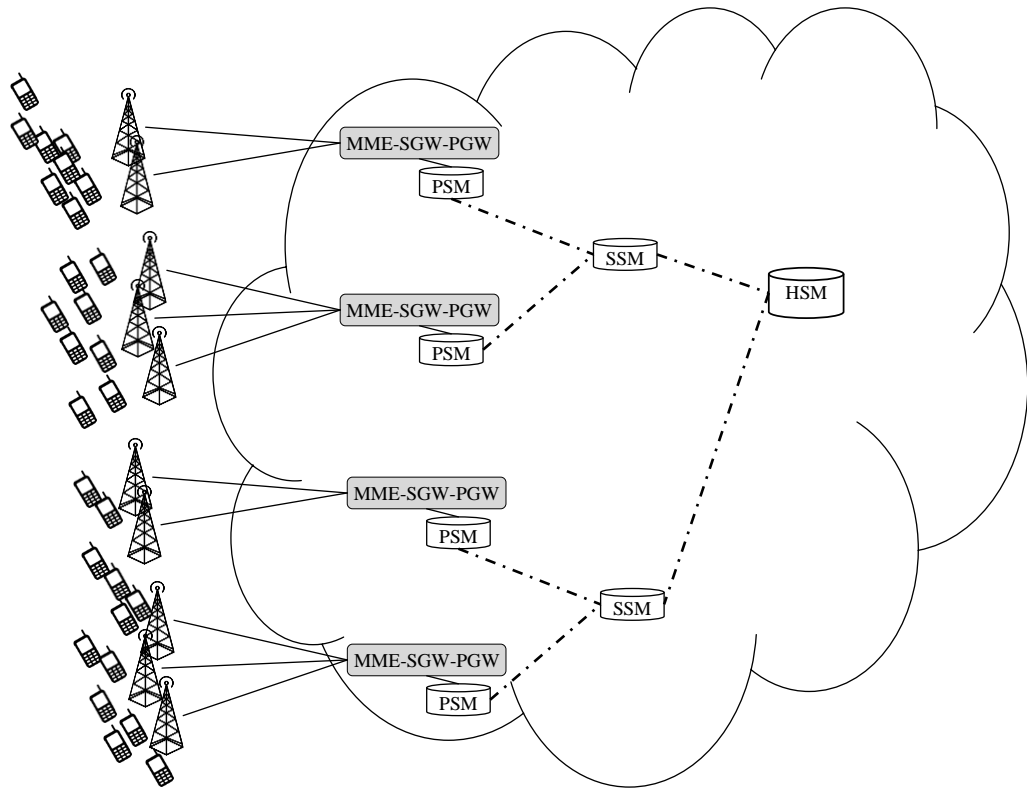


Рисунок 2.13 — Мережна архітектура з системою керування станом

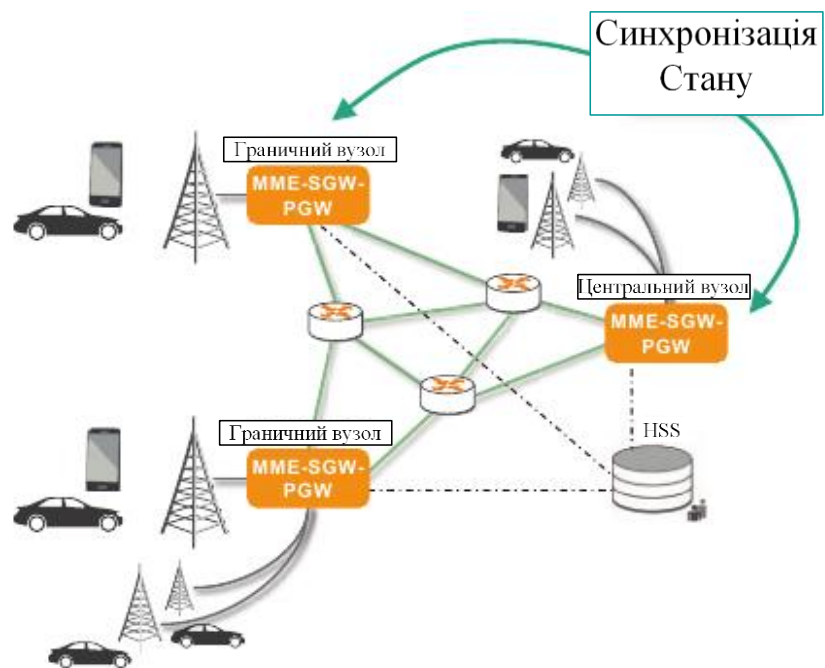


Рисунок 2.14 — Синхронізація стану EPS

Переваги такого підходу різноманітні. По-перше, кожен компонент завжди володіє оновленою інформацією про всіх користувачів і стан синхронізований і узгоджений по всій системі, що дозволяє компонентам бути прозоро доданими або видаленими. Крім того, логіка кожного елементу не знає, що інші компоненти одного і того ж типу знаходяться в системі; що призводить до практично площини управління, що не зберігає стан, в той час як кожен з компонентів відслідковує стан і може бачити повну інформацію про стан в системі. Зосереджуючи увагу на відмовостійкості, можливість легко реплікувати стан в різних надлишкових компонентах дозволяє уникнути ситуації, що стан втрачається в разі відмови компонента і видаляє єдині точки відмови. Нарешті, процедура синхронізації стану відповідає стандарту EPC, вона по своєму дизайну є прозорою до процедур EPC і не вимагає змін у стандартних протоколах зв'язку між компонентами EPC, таким чином, забезпечує швидке впровадження і інтеграцію з існуючими реалізаціями.

Відзначимо, що введення політики балансування і реплікації стану в EN дає можливість будь-якому EN обробляти запити, що надходять від того ж абонента: eNodeB може виконувати операції балансування навантаження з дрібним рівнем та рівнем запитів.

Для того, щоб мати можливість транспортувати інформацію про стан контрольованим чином через мережу, існує потреба в розробці специфічного інтерфейсу і афферентного комунікаційного протоколу. Такий інтерфейс повинен бути визначений в контексті 3GPP, а також в більш загальному контексті, дозволяючи будь-який тип передачі стану.

Далі опишемо Межовий Протокол Синхронізації (Edge Synchronization Protocol – ESP) у вигляді набору різних процедур для координації вузлів EN. Перша процедура називається Процедура Встановлення Інтерфейсу Синхронізації і її основною метою є встановлення початкового загального стану абонента між двома або більше сутностями EN. Друга називається Процедура Повторної Синхронізації, забезпечує належну повторну

синхронізацію двох різних EN, коли зв'язок між ними навмисно або ненавмисно втрачається через різні збережені стани, коли підключення повертається. Протокол може бути використаний між декількома EN в разі стандартних реалізацій EPS 3GPP, а також для будь-якої інформації про стан абонента, яка має формат, аналогічний телекомунікаційному.

2.4.1 Процедура встановлення інтерфейсу синхронізації

Ця операція відбувається, коли новий EN під'єднується до пулу [100]. Припускаючи, що EN1 є вузлом, що приєднується, пропонована процедура синхронізації починається з того, що EN1 звертається з проханням приєднатися до пулу синхронізації стану абонента, де EN2 вже приєднаний. Як показано на рис. 2.15, перша фаза процедури починається з прив'язки між двома об'єктами. EN1 підписується на зміни у інформації стану, з точки зору нової або оновленої, і, навпаки, EN2 робить підписку на модифікації, які відбуваються в EN1 (кроки 1 і 2). Після чого, вони можуть обмінюватися один з одним станами абонентів, які вони в даний час зберігають. Цей крок (крок 3) також виконується щоразу, коли стан оновлюється або приєднується новий абонент, таким чином, після стандартних процедур Attach, Detach та Handover. Крім того, щоб забезпечити узгодженість станів в системі, кожен стан негайно поділяється після завершення процедури.

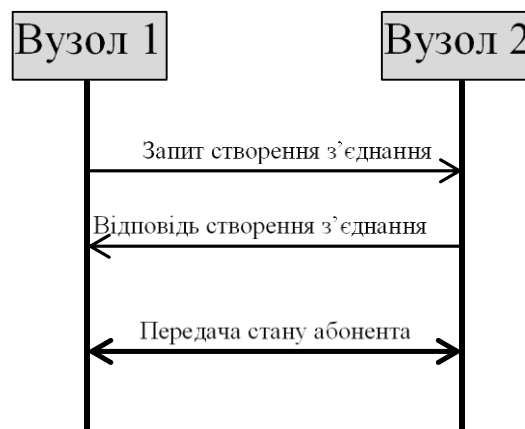


Рисунок 2.15 — Процедура встановлення інтерфейсу сигналізації [100]

2.4.2 Процедура повторної синхронізації

У разі відключення між контролерами, повинна бути виконана процедура синхронізації [100]. Як пояснено нижче, спроектоване рішення засноване на часових мітках, і, таким чином, вимагає синхронізації годинників прилеглих вузлів EN. Використовується Network Time Protocol (NTP), стандартний протокол для синхронізації годинників на мережах з пакетною комутацією та змінної затримкою мереж [101].

Розглядаючи тепер процедуру повторної синхронізації, вона полягає в перевірці обміну часовими мітками вузлів EN кожен раз, коли стан пересилається, в повідомленні «SubscriberStateTransmission», так що під час фази повторного з'єднання кожен EN може знати, який з них володіє новим станом. Приклад описаний на рис. 2.16, в якому під час відключення двох EN, UE модифікує абонентський стан за допомогою процедури, що виконується з EN2 (крок 3). Коли вони стають знову підключені, EN1 посилає стан абонента разом з відповідною міткою часу (TS1) (крок 4). Оскільки TS1 старіше TS2, EN2 відправляє назад новий стан і EN1 оновлює свій власний (кроки 5 і 6).

Підкреслимо, що введені розширення є прозорими для базової мережі 3GPP; іншими словами, описані розширення інтегруються в EPCaaS, підтримуючи його сумісним зі стандартом, в той же час використовуючи прозорі механізми.



Рисунок 2.16 — Процедура повторної синхронізації [100]

2.5 Оркестрування віртуалізованої мережі EPC

Загальна архітектура системи керування гібридною мережею представлена на рис. 2.17.

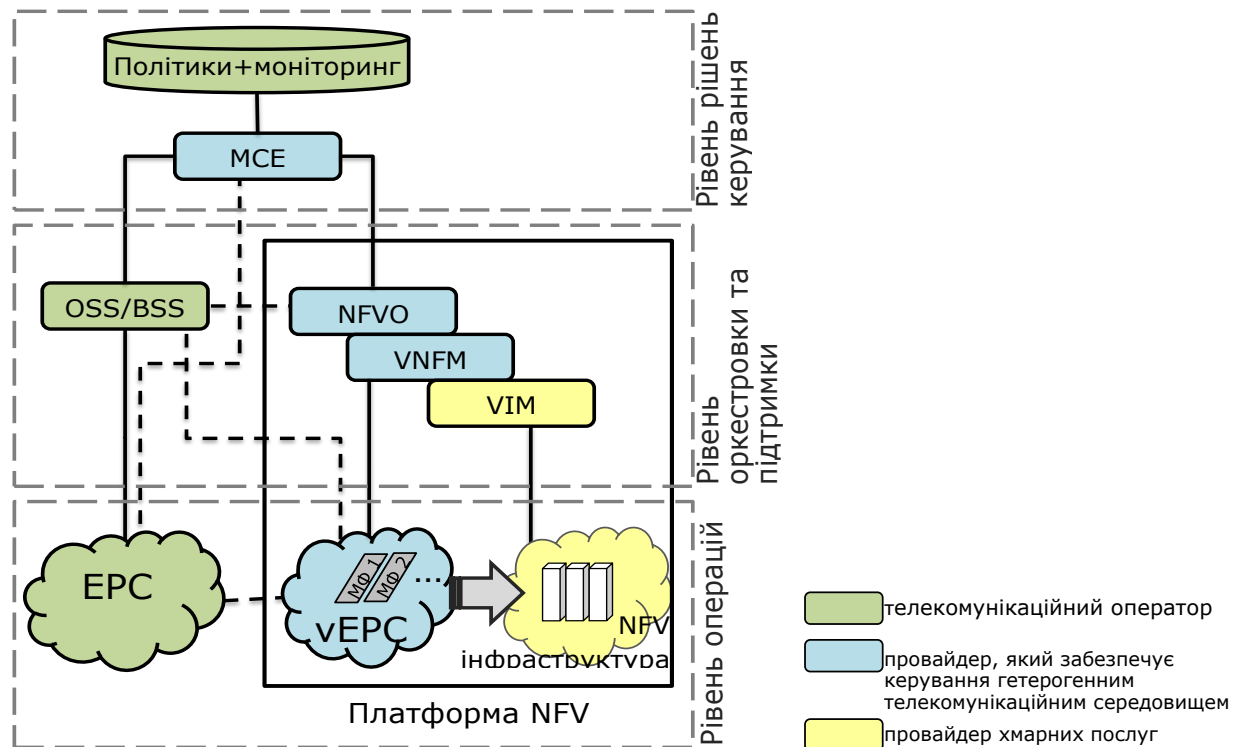


Рисунок 2.17 — Архітектура системи керування гібридною мережею

Процедура створення віртуалізованої базової мережі (vEPC) за запитом починається з того, що подія моніторингу посилається від традиційної мережі LTE EPC (LEPC) у напрямку до OSS/BSS (крок 1), як показано на рис. 2.18 [102].

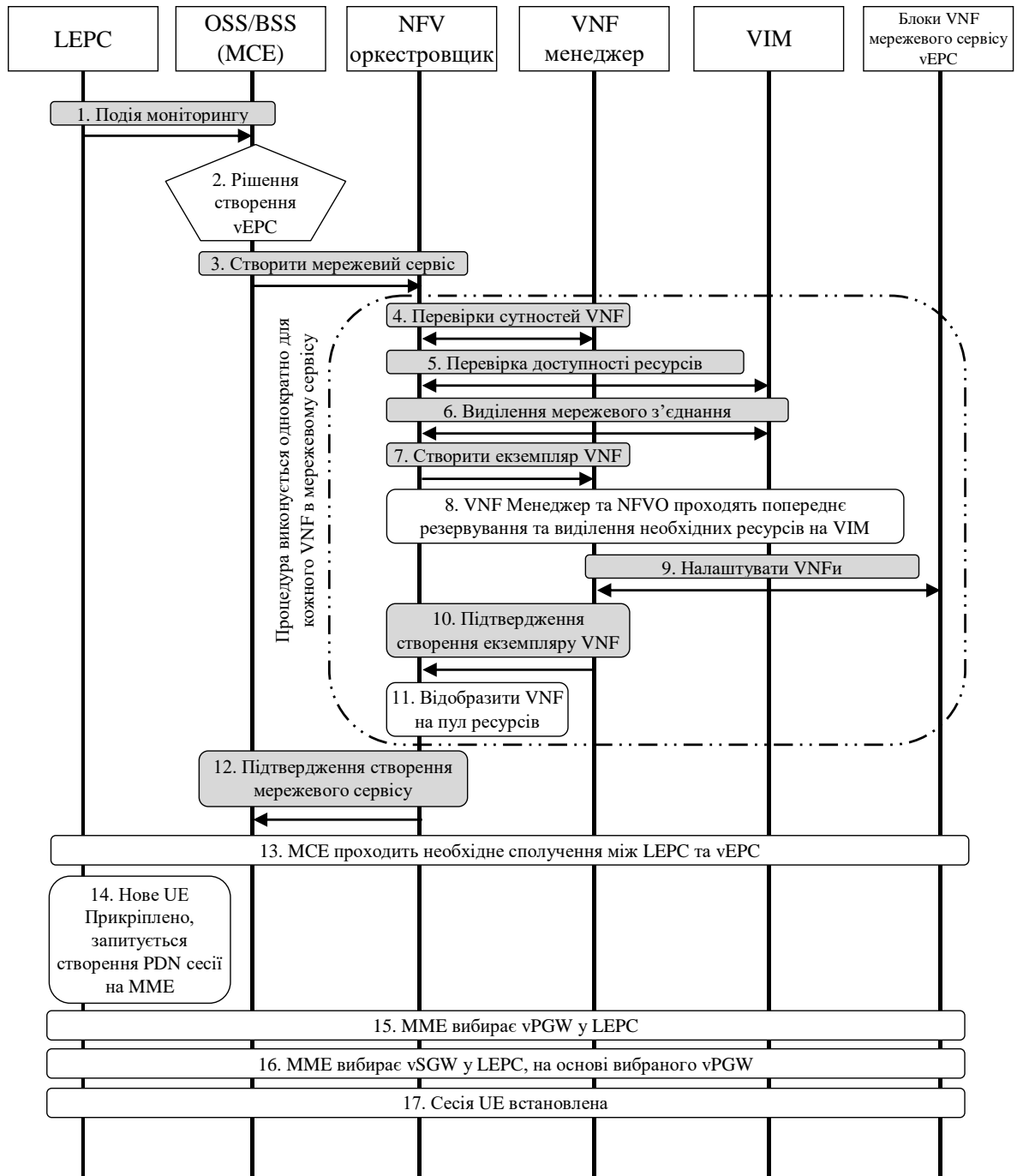


Рисунок 2.18 — Потік повідомлень створення сутностей VNF для vEPC

Цей приклад ілюструє специфічним випадок режиму моніторингу OSS/BSS, де сутності здатні збирати метрики мобільних мереж від різних

елементів EPC, оцінюючи операційний стан мережі і повідомляючи його в блок керування Management Component Entity (MCE). За допомогою цієї інформації, MCE (інтегрований в OSS/BSS для спрощення рисунку) може прийняти рішення про створення сутностей віртуалізованого мережевого сервісу EPC, дозволяючи оператору вирішувати вимоги зростання сервісу (крок 2). Таким чином, команда створення сутностей відправляється до NFVO (крок 3). Це ініціює процедуру, яка повторюється для кожного VNF, що складає NS (що формується через граф передачі VNF), що починається з перевірки NFVO існуючих екземплярів VNF в VNFM (крок 4). Це дозволяє йому визначити, які нові VNF повинні бути створені і для кожного VNF NFVO запитує VIM наявні ресурси (крок 5), а також запитує виділення мережових з'єднань між VNF, які є частиною мережевого сервісу, над віртуалізованим мережовим ресурсом (крок 6). Після завершення попередніх етапів, умови збираються NFVO для запиту фактичного створення екземплярів сутностей окремих VNF до VNFM (крок 7), що охоплює попереднє резервування і виділення необхідних ресурсів на VIM (крок 8). VNFM переходить до конфігурації конкретних VNF, відповідно до їх специфікації дескрипторів або конкретних параметрів розгортання (крок 9). У разі якщо VNF вимагає сполучення з блоком Управління Елементом (Element Management), цей крок конфігурації дозволяє VNFM повідомити ЕМ, який потім може застосувати параметри специфічні для прикладної програми до VNF після їх створення. VNFM потім відповідає назад NFVO про їх успішну фіналізацію створення сутностей (крок 10), і далі відображення на пул ресурсів (крок 11). Потім процедура підтверджує наявність мережевого сервісу зі сторони MCE (крок 12), який потім може приступити до необхідного сполучення між фізичними елементами LEPC, і їх віртуалізованими відповідностями, які знаходяться в vEPC (крок 13). Тобто, HSS, PCRF та MME інформуються про створену vEPC. Таким чином, відповідні інформаційні контексти стосовно наявних vPGW та vSGW оновлюються в цих сутностях.

Коли UE приєднується до LEPC (крок 14), запит на приєднання UE обробляється в MME, який виконує вибір PGW та SGW для приєднаного UE, виконуючи стандартну процедуру 3GPP. Передбачається, що вагові коефіцієнти IP-адрес vPGW/vSGW, повернені з DNS-сервера, перевищують значення інших IP-адрес PGW/SGW. Для цього, як тільки мережа vEPC буде створена і готова до сервісу vEPC, нові DNS записи створених vPGW/vSGW повинні бути додані до внутрішнього DNS OSS/BSS.

Приклад робочого процесу оркестрації сервісу представлено на рис. 2.19.

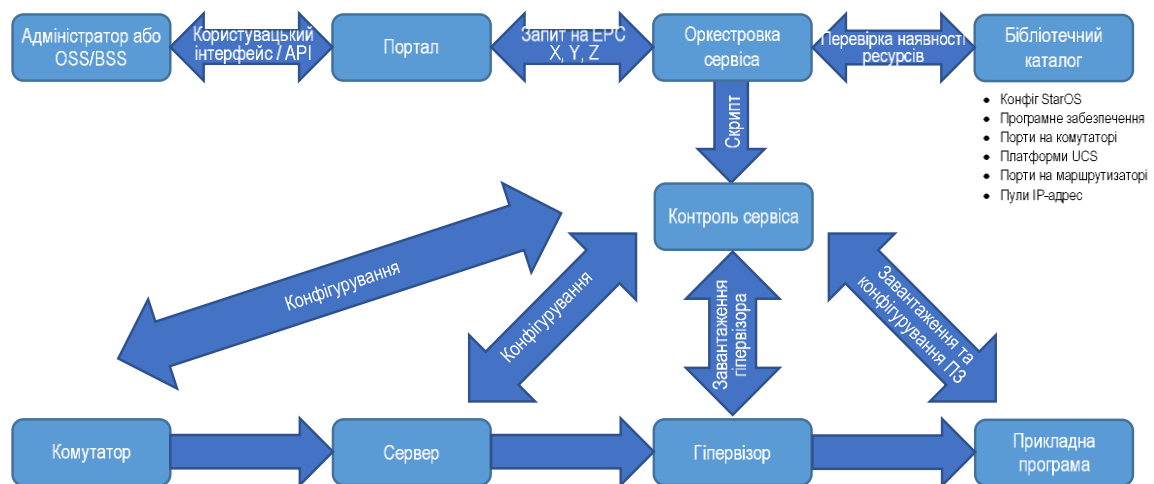


Рисунок 2.19 — Приклад робочого процесу оркестрації сервісу [103]

2.6 Оцінка методу визначення місця розміщення та необхідної ємності віртуальних ресурсів у разі виникнення перевантаження фізичної мережі

Вирішення задачі (2.13)-(2.22) здійснювалось з використанням генетичного алгоритму з дійсним кодуванням у системі MATLAB. Фрагмент коду програми представлено на рис. 2.20.

```

>>ObjectiveFunction=@fitness;
>>nvars=108;% Number of variables
>>LB=zeros(1,108);% Lower bound
>>UB=ones(1,90);% Upper bound
>>IntCon=ones(1,90);% Integer variables
>>ConstraintFunction=@constraint;
>>[x,fval]=ga(ObjectiveFunction,nvars,[],[],[],[],LB,
UB,ConstraintFunction,IntCon,options)

```

Рисунок 2.20 — Фрагмент коду MATLAB

Приклад моделювання системи з десятима вузлами, трьома функціональними блоками та двома типами ресурсів показав зменшення умовних витрат в середньому до 15% при використанні методу визначення оптимального місця розміщення ресурсів у порівнянні зі стратегією коли ЕРС як сервіс надається єдиним наперед визначеним хмарним вузлом (рис. 2.21).

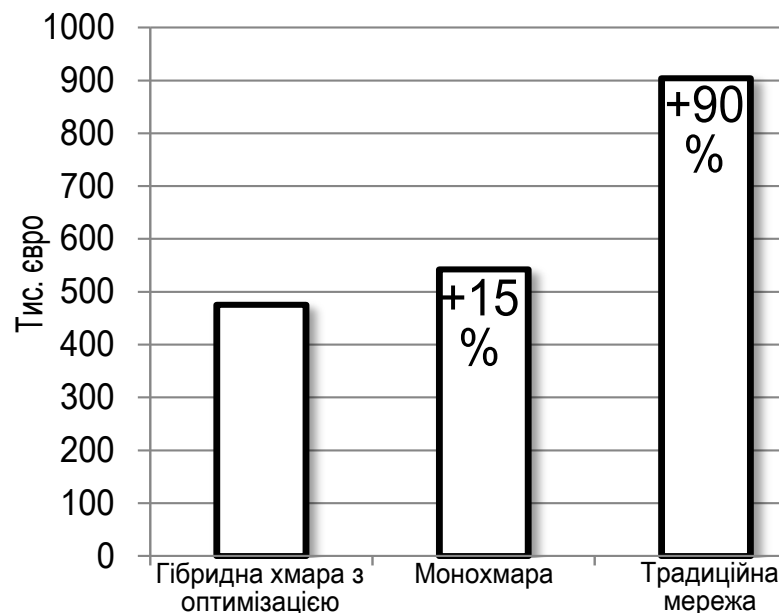


Рисунок 2.21 — Загальна вартість розміщення

На рис. 2.22 показано приклад знайдених значень кількості зарезервованих ресурсів за методом (2.13)-(2.22).

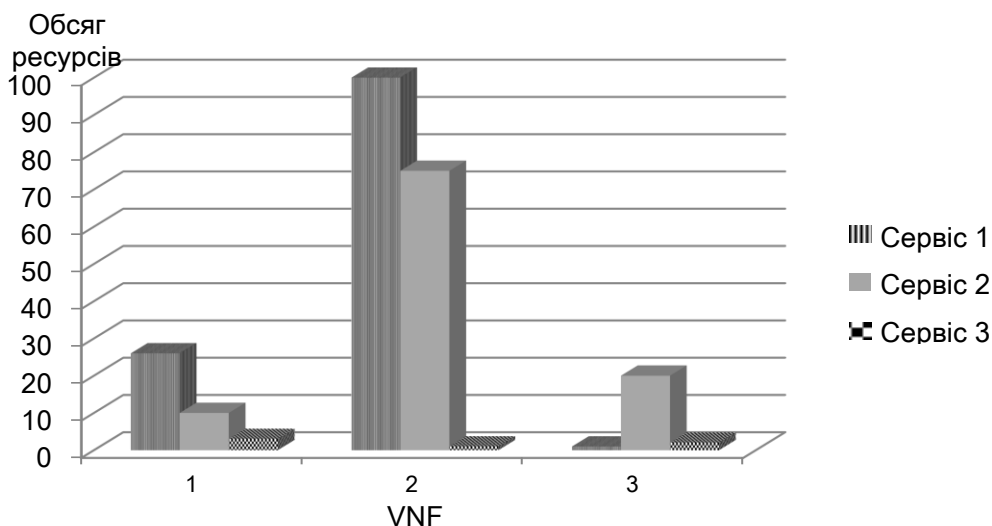


Рисунок 2.22 — Обсяги зарезервованих ресурсів для мережевих функцій EPC

В роботах [104-111] представлено основні результати розділу, де здобувачем запропоновано системний підхід до організації обчислювально-телекомунікаційного середовища обслуговування службового трафіку мобільних абонентів.

Висновки

1. Запропоновано метод керування розміщенням та визначенням необхідного обсягу ресурсів мережевих функцій для оптимізації кількості ресурсів, виділених мережевій функції у мережі телекомунікаційного оператора.

2. Запропонований метод дозволяє пов'язати продуктивність мережевої віртуалізованої функції з виділеною їй кількістю ресурсів, а також враховує гібридність телекомунікаційного середовища.

3. Розглянуто архітектурне рішення еластичного керування EPC за допомогою поділу стану, а також процес взаємодії між мережними елементами від час розгортання віртуалізованих мережних вузлів.

4. Метод може застосовуватись при керуванні розгортанням мережевих функцій у гібридному апаратному середовищі для мінімізації витрат оператора зв'язку та покращення якості обслуговування абонентів.

РОЗДІЛ 3

СИСТЕМА КЕРУВАННЯ РЕСУРСАМИ ВІРТУАЛІЗОВАНИХ МЕРЕЖЕВИХ ФУНКЦІЙ

3.1 Постановка задачі

Метою системи керування ресурсами мережеских функцій є виділення достатньої їх для задоволення вимог до якості обслуговування навантаження. В основі будь-якого алгоритму надання ресурсів лежать два питання: скільки надавати і коли [78].

Скільки надавати: Для вирішення питання про те, скільки ресурсів виділити для кожної мережескої функції, використовуємо модель серверу з розділенням процесору. Модель приймає в якості вхідних даних інтенсивність надходження вхідних запитів і вимоги обслуговування окремого запиту, і обчислює кількість ресурсів, необхідних мережескій функції, щоб впоратися з вимогами.

Коли надавати: Рішення про те, коли надавати ресурси, залежить від динаміки навантажень. Телекомунікаційні навантаження зазнають довгострокових змін, таких як вплив години дня або сезонні ефекти, а також короткострокових коливань таких як несподівані натовпи. У той час як довгострокові коливання можуть бути передбачені заздалегідь, спостерігаючи за змінами в минулому, короткострокові коливання менш передбачувані, а в деяких випадках, не передбачувані. Запропонована методика використовує два різних методи для роботи в умовах змін, які спостерігаються в різних часових масштабах. Використовується прогностичне управління ресурсами для оцінки навантаження і відповідного керування, а також реактивне керування ресурсами для реагування на непередбачені зростання у кількості заявок, що надходять.

Розглянемо мережу, в якій функціонує кілька мережеских функцій [67]. Передбачається, що кожна така мережеска функція вказує бажану вимогу до

якості обслуговування (QoS); при цьому в даному випадку передбачаємо, що вимоги до QoS визначені в термінах цільового часу відповіді. Метою системи є забезпечення того, що середній час відповіді (або деякий процентиль часу відповіді), який спостерігається запитами мережевої функції не перевищує бажаний цільовий час відповіді. Загалом, кожен вхідний запит обслуговується декількома апаратними та програмними ресурсами на сервері, такими як CPU, NIC, диск і т.д. Припускаємо, що заданий цільовий час відповіді розділяється на кілька значень часу відповіді для конкретних ресурсів по одному для кожного такого ресурсу. Таким чином, якщо кожен запит на кожному ресурсі не витрачає часу більше, ніж призначене цільове значення, то загальний цільовий час відповіді для сервера буде задоволений. Задача розділення зазначеного значення часу відповіді сервера на значення часу відповіді для конкретного ресурсу виходить за рамки даного дослідження; у роботі передбачається, що такі конкретні для ресурсу значення часу відповіді задані.

Системи обслуговування, що представляють собою ресурси (процесор та пам'ять), передбачаються послідовними, тобто обслуговування протікає конвеєром. Додатково, приймається дисципліна обслуговування з розділенням процесора (processor sharing – PS) у кожній черзі, оскільки PS апроксимує політику диспетчеризації, що використовується більшістю операційних систем, наприклад, зважену справедливую організацію черги та розподіл машинного часу у Linux. Для простоти подальшого викладу припускаємо, що в системі наявний лише один тип ресурсу.

Формально, позначимо цільовий час відповіді мережевої функції i як P_i і T_i – спостережуваний середній час відповіді, тоді мережеві функції потрібно виділити таку кількість ресурсів, щоб $T_i \leq P_i$.

Використаємо таку постановку задачі щоб одержати механізм динамічного виділення ресурсів, який описано далі.

3.2 Система динамічного керування ресурсами

Щоб виконати динамічне виділення ресурсів, засноване на вищевказаному формулюванні, на кожному сервері необхідно буде використовувати три компоненти: (1) модуль моніторингу, який вимірює навантаження і показники продуктивності кожної мережевої функції (такі як інтенсивність надходження запитів, середній час відповіді тощо), (2) модуль прогнозування, який використовує вимірювання з модуля моніторингу для оцінки характеристик навантаження в найближчому майбутньому, і (3) модуль розподілення ресурсів, який використовує ці оцінки навантаження для визначення кількості ресурсів, яку необхідно виділити мережевим функціям. На рис. 3.1 показані ці три компоненти [67].

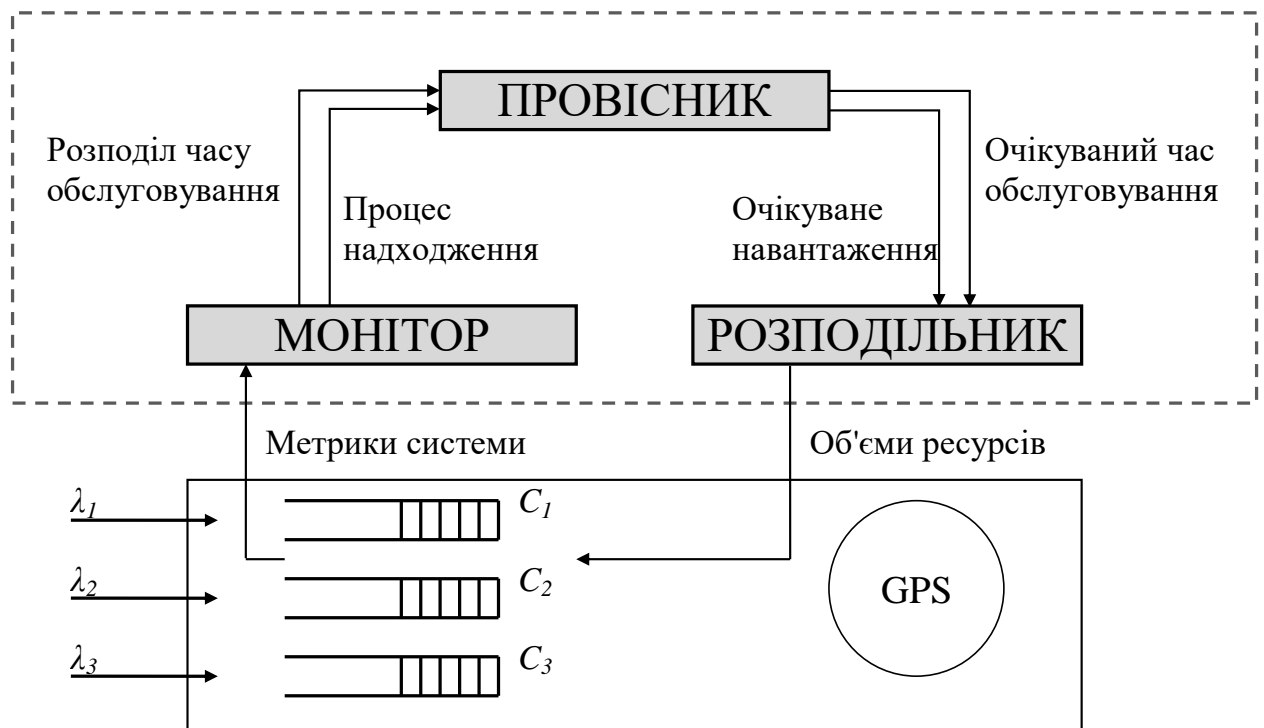


Рисунок 3.1 — Динамічне виділення ресурсів

Як вже відмічалось, навантаження часто мають передбачувані шаблони. Проте, можуть бути значні відхилення від цих моделей через пульсуючий характер навантажень. Надлишкові заявки або раптові сплески запитів можуть привести до погіршення обслуговування. Крім того, існують витрати

та ризики пов'язані зі змінами у системі, і, таким чином, прагнуть уникнути частих змін конфігурації ресурсів. Ґрунтуючись на цих спостереженнях, пропонується гібридне рішення системи керування виділенням ресурсів, яке поєднує в собі прогностичну складову з реактивною.

На рис. 3.2 показана концептуальна архітектура пропонованого рішення:

1. Провісник базового навантаження аналізує історичні дані навантаження і визначає закономірності, які формують базове навантаження.
2. Координатор направляє запити навантаження на сервери, а також обмінюється даними з контролером для надання інформації про вхідне навантаження.
3. Контролер оцінює і виділяє відповідну кількість ресурсів необхідну для обробки навантаження. Він генерує розподіл ресурсів (кількість ресурсів на рис. 3.2).

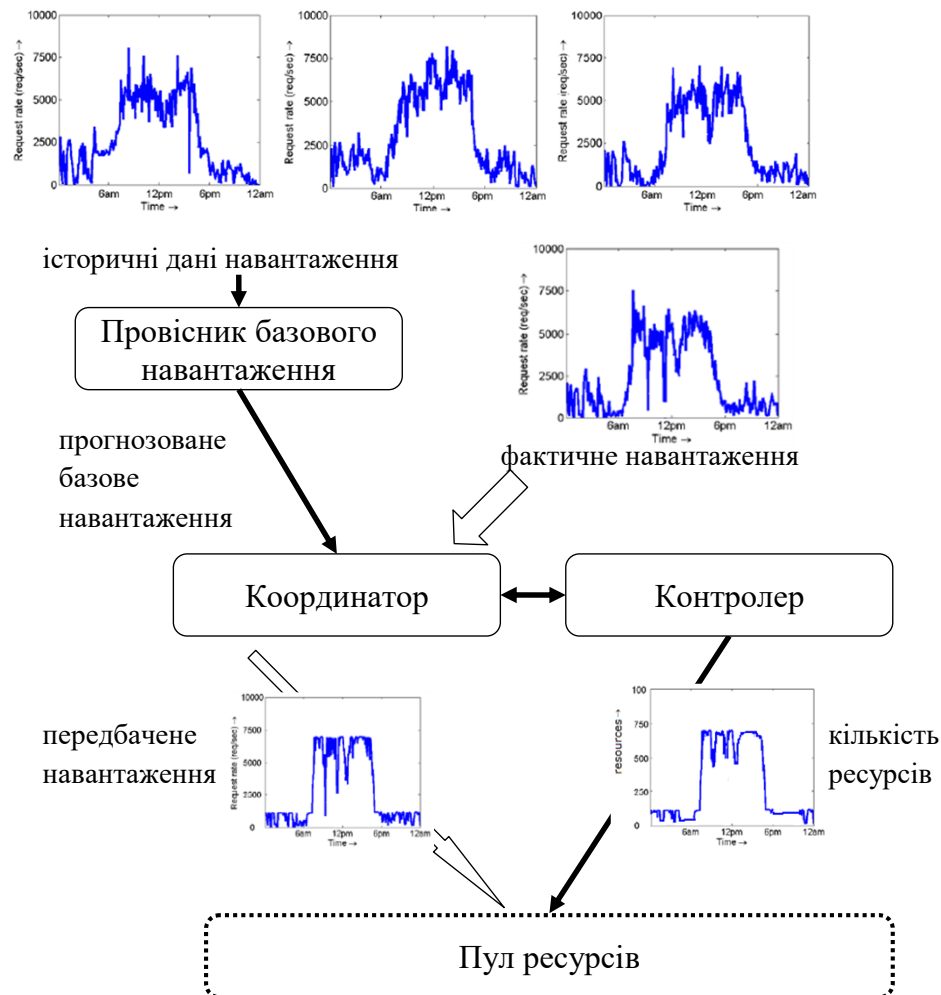


Рисунок 3.2 — Система гібридного керування

3.3 Моніторинг та керування

Модуль онлайн моніторингу відповідає за вимірювання різних метрик системи і мережевих функцій. Ці показники використовуються для оцінки параметрів моделі системи і характеристик навантаження.

Моніторинг на мережі реалізується встановленням головної прикладної програми моніторингу та відповідних підконтрольних їй агентів на кожен мережевий блок [79].

Важливий аспект платформи полягає в здатності моніторингу її різних складових елементів [102]. Цей механізм дозволяє платформі адаптувати і масштабувати мережеві функції у відповідності з різними метриками, де сутності прийняття рішень порівнюють отримані тригери з вимогами описів сервісів окремих елементів, які повідомляються ним. Платформа забезпечує можливість відділення аспектів моніторингу функціонування мережі мобільного зв'язку від тих, які відносяться до інфраструктури NFV, підтримуючи існуючі механізми керування та оркестровки в зв'язку з цим. Конкретно, інтерпретація подій моніторингу OSS/BSS і MCE з фізичної LTE EPC може бути використана для запуску створення віртуальної EPC (vEPC) в інфраструктурі NFV, або навіть просто створення одного елемента vEPC або компонента. Таким чином, доступні механізми моніторингу в платформі надаються на різних рівнях:

- VNF: коли VNF вбудовує функції моніторингу і передає зібрану інформацію в Element Management (EM) або VNF Manager (VNFM). Система дозволяє функціям VNF безпосередньо направляти дані сутностям OSS/BSS і MCE з подіями моніторингу, що дозволяє їм приймати загальномережеві управлінські рішення, з урахуванням також віртуалізованих функцій;
- EM: EM може реалізувати рішення про дії, на основі даних, зібраних з VNF, і направити їх на VNFM, або до OSS/BSS і MCE для дій на всю мережу;

- OSS/BSS і MCE: коли функція моніторингу не реалізована в ЕМ, або перетинає кілька ЕМ. Це може бути як точкою детектування, так і рішення. OSS/BSS також отримує події моніторингу LTE EPC, що дозволяє запускати зміни в віртуалізованій мережі мобільного зв'язку або її елементів;
- Оркестровщик NFV (NFVO): коли робота віртуалізованого сервісу зазнає впливу на рівні мережевого сервісу (тобто складається з декількох взаємопов'язаних VNF з загальним кінцем), то NFVO може бути покращено можливістю спільно розглядати різні звіти про моніторинг окремих VNF, і діяти, коли функціональні потреби мережевого сервісу в цілому піддаються впливу, запускаючи VNFM діяти над необхідними VNF;
- VNFM: він приймає рішення масштабування, основуючись на подіях VNF, відповідно до параметрів масштабування і моніторингу в їх дескрипторах. Цей об'єкт може використовувати цю інформацію, а також діяти у відношенні необхідних дій інфраструктури NFV, отримуючи події від менеджера віртуалізованої інфраструктури (Virtualized Infrastructure Manager – VIM).

Таким чином, результат моніторингу в розглянутій платформі спрямований в бік загальної поведінки масштабування віртуальних ресурсів відповідно до продуктивності мережевого сервісу/VNF, трьома шляхами: 1) автоматичне масштабування, де VNFM моніторить події і запускає операцію масштабування при виконанні певних умов, 2) масштабування на вимогу, коли VNF, або його ЕМ, викликає операцію масштабування за допомогою явного запиту на VNFM, або 3) запит керування, який ініціюється відправником (OSS/BSS або MCE) у напрямку до VNFM через NFVO.

Для того, щоб здійснювати моніторинг роботи площини користувача і площини управління, S1-U і S5 є точками спостереження для моніторингу продуктивності трафіку даних, в той час як S1-MME і S10 є точками спостереження для моніторингу продуктивності навантаження сигналізації через необхідні операції і активність абонентів для користувальницького

обладнання. Інформація моніторингу з означених опорних точок спостереження збирається на OSS/BSS. Інформація від сутностей компонентів vEPC збирається на кожному ЕМ і доставляється зрештою OSS/BSS. Система дозволяє мережевим функціям безпосередньо направляти дані сутностям OSS/BSS і MCE з подіями моніторингу, що дозволяє їм приймати загальномереві управлінські рішення, з урахуванням також віртуалізованих функцій.

Диференціація інтенсивності моніторингу окремих сегментів мережі [79] дає можливість усунути надлишкові процеси моніторингу та обробки даних, забезпечуючи необхідну інтенсивність лише на тих вузлах, де це необхідно. Використовуючи такий підхід можливо підвищити гнучкість управління мережевими функціями, і, як наслідок, ефективніше використовувати ресурси пристроїв рівня управління. Крім того, володіючи актуальною службовою інформацією, можна здійснювати оптимальне керування навантаженням у мережі. Це дозволить значно зменшити витрати на електроенергію та збільшити час «життя» елементів мережі.

3.4 Визначення інтервалу часу сталої конфігурації ресурсів

Оскільки мережевий трафік не є статичним це може привести до неефективного використання загальних ресурсів протягом дня, з одного боку. З іншого, необхідно забезпечувати задану якість обслуговування. Тому пропонується стратегія коли конфігурація віртуальних машин створюється завчасно з розгляду найгіршого випадку, а активізується коли виникає необхідність.

Пропонується застосовувати метод динамічного визначення розміру інтервалу часу сталої конфігурації ресурсів, тобто визначення періодів часу протягом яких конфігурація ресурсів буде незмінною – часу між двома послідовними запусками алгоритму реконфігурації. Суть цього механізму полягає у динамічній зміні інтенсивності здійснення перерозподілу ресурсів

залежно від різниці між мінімальним та максимальним значенням навантаження мережевої функції на інтервалі на основі зібраних за довготривалий період часу даних стану елемента мережі. Якщо це значення збільшується, то скорочується інтервал сталої конфігурації мережевих функцій.

Формула (3.1) описує принцип зміни інтервалу часу сталої конфігурації ресурсів:

$$Int(t) = \max \left(Int_{base} \cdot \left(1 - K \cdot \frac{\max_{\tau \in (t-I(t-1);t)} \lambda_{basepred}(\tau) - \min_{\tau \in (t-I(t-1);t)} \lambda_{basepred}(\tau)}{\max \lambda_{basepred}} \right); Int_{minbase} \right), \quad (2.1)$$

де Int – інтервал часу сталої конфігурації ресурсів, Int_{base} – базове значення інтервалу, $Int_{minbase}$ – мінімальна допустима величина базового інтервалу, K – коефіцієнт зміни тривалості сталої конфігурації, $\lambda_{basepred}(t)$ – середньостатистична передбачена інтенсивність надходження заявок під час інтервалу t .

Визначаємо і дискретизуємо шаблони у прогнозованому навантаженні. Мета полягає в тому, щоб представити денний шаблон в навантаженні дискретизуючи його запити у послідовні, непересічні часові інтервали з єдиним репрезентативним значенням вимоги в кожному інтервалі. Запропоновано алгоритм для знаходження невеликої кількості часових інтервалів, таких, що відхилення від фактичної вимоги зведено до мінімуму. Підтримувати невелику кількість інтервалів важливо, так як більше число інтервалів означає більш часті зміни виділення ресурсів і, таким чином, більш високі ризик і витрати. Визначимо формально дискретизацію [51].

Дискретизація навантаження: Маючи часовий ряд X на області $[v, \tau]$, часовий ряд Y на тій же самій області є характеристикою/дискретизацією навантаження X якщо $[v, \tau]$ може бути розділена на m послідовних

непересічних часових інтервалів, $\{[\tau_1, \tau_2], [\tau_2, \tau_3], \dots, [\tau_{m-1}, \tau_m]\}$, так що $X(j)=r_i$, для всіх j у i -му інтервалі, $[\tau_{i-1}, \tau_i]$.

Зверніть увагу, за визначенням, будь-який часовий ряд, X , є дискретизацією сам по собі. Встановлюємо $v=0$ і τ =період навантаження. У подальшій дискусії припускаємо період у 24 години.

Ідея дискретизації подвійна. По-перше, потрібно точно відображати вимоги. Для досягнення цієї мети, репрезентативні значення, r_i , для кожного інтервалу, $[\tau_{i-1}, \tau_i]$, мають бути якомога ближчими до фактичних значень часового ряду в інтервалі $[\tau_{i-1}, \tau_i]$. По-друге, виділення ІТ-ресурсів не відбувається безоплатно [8]. З цієї причини, потрібно уникати занадто великої кількості інтервалів і, отже, занадто великої кількості змін в системі, так як це не практично і може призвести до багатьох проблем (наприклад, втрати продуктивності, зносу серверів, нестабільності системи і т.д.), Таким чином, слід мінімізувати помилку, внесену дискретизацією, і кількість інтервалів в дискретизації. Пропонується рішення для дискретизації часових рядів таким чином, щоб зменшити обидві величини.

$$w_1 \cdot f_1(m) + w_2 \cdot f_2(m) \rightarrow \min \quad (3.2)$$

Вираз (3.2) є цільовою функцією яку хочемо мінімізувати, де w_1, w_2 – коефіцієнти нормалізації, $f_1(m)$ є функцією відхилення репрезентації від кривої навантаження при m інтервалах (наприклад, лінійна), $f_2(m)$ є функцією вартості кількості змін (інтервалів m).

$$f_1(m) = \sum_{i=1}^m \left[\sum_{\tau=\tau_{i-1}}^{\tau_i} u \left(r_i - \lambda_{basepred}(\tau) \right) \right] \quad (3.3)$$

$$r_i = \max_{\tau \in (\tau_{i-1}, \tau_i)} \lambda_{basepred}(\tau) \quad (3.4)$$

Метою виразу (3.2) є одночасна мінімізація помилки репрезентації навантаження і кількості змін.

У деяких випадках, можна було б віддати перевагу, мінімізації квадрату кількості змін (або будь-якій іншій функції кількості змін). Вибір кращої функції вартості повинен здійснюватися мережевими адміністраторами з урахуванням конкретних потреб мережі під управлінням [112]. Для даної роботи встановлюємо $f_2(m)=l*m$, де l є константою нормалізації. Цільова функція виражає мету, мінімізуючи нормалізовану кількість змін і помилку репрезентації. Функція вартості репрезентативної помилки використовується для кількісної оцінки похибки подання навантаження. Зверніть увагу, що в найзагальнішому випадку, як кількість змін, так і репрезентативна помилка можуть бути сформульовані як функції корисності.

Функція вартості кількості змін, наприклад, може відображати вартість електроенергії, а також грошову вартість пов'язану зі зносом на перезавантаження, яка базується приблизно на вартості заміни диска і номінальному часі напрацювання на відмову [113].

Оптимальну величину базового інтервалу визначасмо задаючи різні значення кількості інтервалів та обчислюючи значення виразу (3.2) і обираючи найкраще, тобто мінімальне, при цьому:

$$Int_{base} = \frac{\tau_m}{m} \quad (3.5)$$

Псевдокод алгоритму (рис. 3.3) буде мати вигляд, представлений на рис. 3.4.



Рисунок 3.3 — Визначення базового інтервалу

```

interval ← maxinterval
iinterval ← 0
startpoint ← begining
while interval > 0
  startpoint ← 0
  endpoint ← startpoint + interval
  while startpoint < ending
    maxvalue ← max[Load(startpoint)... Load(endpoint)]
    for j ∈ startpoint... endpoint
      BaseLoadinterval(j) ← maxvalue
      DiffLoadinterval(j) ← BaseLoadinterval(j) - Load(j)
    startpoint ← endpoint
    endpoint ← startpoint + interval
  iinterval ← iinterval + 1
  interval ← interval - 1
for j ∈ 1...maxinterval
  Minfuncj ← ∑k DiffLoadj(k) + l * ij
MinInt ← argmin(Minfunc)
MinCost ← min(Minfunc)
  
```

Рисунок 3.4 — Псевдокод алгоритму визначення оптимального базового інтервалу

З іншого боку, можемо додатково накласти умову, що різниця між сусідніми репрезентативними значеннями навантаження не повинна бути меншою певного заданого порогу. Тобто для всіх інтервалів проводиться аналіз r_i , якщо $|r_i - r_{i+1}| < \varepsilon$, де ε – деякий поріг, тоді значення i та $(i+1)$ об'єднуються, а нове репрезентативне значення обирається як $\max(r_i, r_{i+1})$.

3.5 Розподілення ресурсів мережевих функцій

Опишемо модель, яка використовується для визначення потреби в ресурсах мережевої функції на основі очікуваного нею навантаження і цільового часу відповіді.

Оскільки ресурс моделюється як GPS сервер, інтенсивність обслуговування запиту мережевої функції дорівнює d/s , де d – це кількість ресурсів мережевої функції і s – це середній час обслуговування запиту однією одиницею ресурсу. Отже, інтенсивність обслуговування запиту

$$\mu = d/s. \quad (3.6)$$

Засновуючись на теорії масового обслуговування, розподіл часу обслуговування заявок такої системи є експоненційними з середньою величиною $\frac{1}{\mu - \lambda_{basepred}}$. Мережевій функції потрібно виділити кількість ресурсів, так що $T \leq P$. Цей вираз можна переписати як

$$T = \frac{1}{\mu - \lambda_{basepred}} \rightarrow d_i \geq s \left(\frac{1}{T} + \lambda_{basepred} \right). \quad (3.7)$$

3.6 Метод прогнозування навантаження

Визначення конфігурації ресурсів описане в попередньому пункті у значній мірі залежить від точної оцінки навантаження, що може надійти.

Пропонується здійснювати постійний моніторинг значень інтенсивності навантаження та використовувати прогностичний метод, якщо ці значення не перевищують порогові, в іншому випадку – оцінювати поточні тренди у навантаженні та здійснювати масштабування ресурсів на основі нового прогнозу.

У цьому пункті представляємо метод прогнозування у разі несподіваних відхилень від початкового прогнозу, що використовує минулі спостереження для оцінки майбутнього навантаження мережевої функцій.

Найважливішим параметром, який характеризує вхідне навантаження, є інтенсивність надходження запитів, і точне її визначення дозволяє оцінювати середній час перебування заявки у системі.

Провісник навантаження використовує минулі спостереження за навантаженням, щоб передбачити пікову вимогу, яка буде зазнаватися протягом часу I . Для простоти викладу припустимо, що $I = 1$ година. Щоб зробити це модуль прогнозування зберігає історію інтенсивності надходження заявок, яка мала місце протягом кожної години дня, протягом останніх декількох днів. Потім для кожної години генерується гістограма, використовуючи спостереження для цієї години за останні декілька днів (див. рис. 3.5). Кожна гістограма дає розподіл ймовірностей інтенсивності надходження за цю годину. Пікове навантаження для певної години оцінюється як високий процентиль розподілу інтенсивності надходження за цю годину (див. рис. 3.5). Таким чином, використовуючи хвіст розподілу інтенсивності надходження для передбачення пікової вимоги, можна визначити достатній обсяг ресурсів, щоб впоратися з навантаженням найгіршого випадку, якщо воно надійде. На додаток до використання спостережень за попередні дні, для прогнозування використовуються навантаження в останні кілька годин поточного дня.

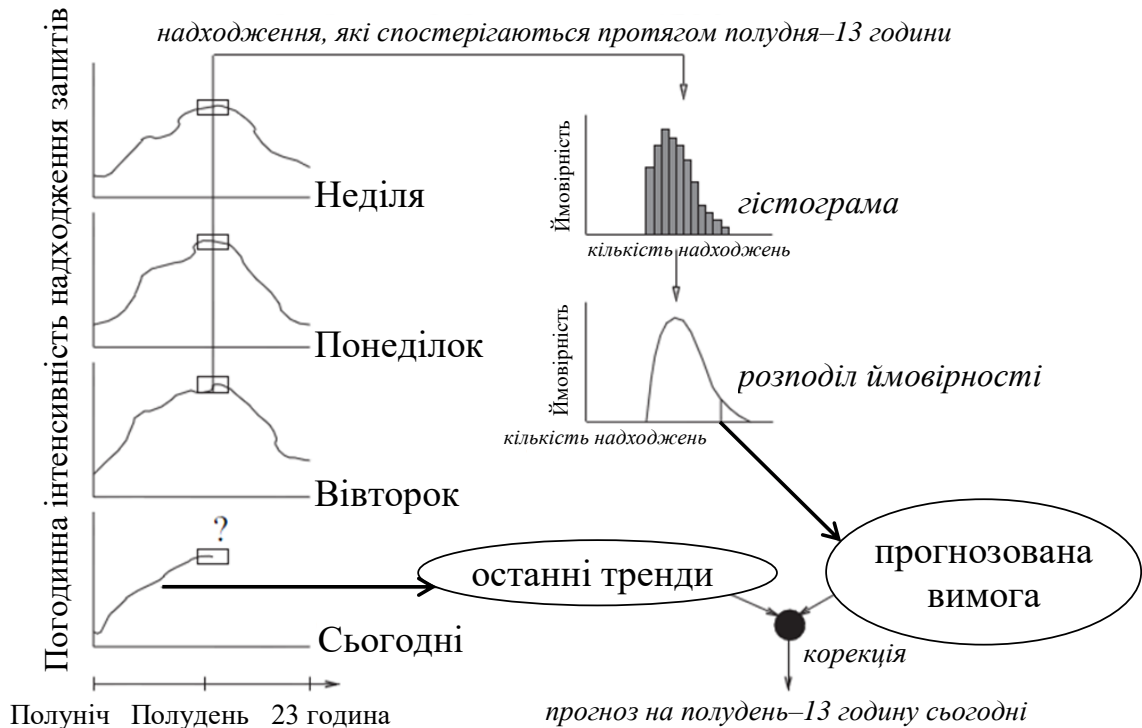


Рисунок 3.5 — Схема передбачення навантаження

Таким чином, пропонується здійснювати прогнозування навантаження на наступний інтервал часу шляхом урахування прогнозу на основі довгострокової статистики та коригування його за моделлю експоненційного згладжування (exponentially weighted moving average) [114], де помилки більш нових минулих періодів мають більший ваговий коефіцієнт. α (константа згладжування) – коефіцієнт, що характеризує швидкість зменшення ваг, та приймає значення від 0 до 1, чим значення цього параметра ближче до одиниці, тим більше при прогнозі враховується вплив останніх рівнів ряду. Параметри моделі встановлюються адміністратором мережі відповідно до експерименту.

Припустимо, що $\lambda_{basepred}(t)$ – передбачена інтенсивність надходження під час певного інтервалу позначеного t . Далі, нехай $\lambda_{obs}(j)$ – реальна інтенсивність надходження протягом інтервалу j . Помилка передбачення це $(\lambda_{obs}(j) - \lambda_{basepred}(j))$. У випадку систематичної помилки передбачення протягом останніх декількох інтервалів, прогнозоване значення протягом

наступного інтервалу коригується з використанням помилки, що спостерігається:

$$\lambda_{pred}(t) = \lambda_{basepred}(t) + \alpha \sum_{j=t-h}^{t-1} (1 - \alpha)^{t-1-j} \cdot (\lambda_{obs}(j) - \lambda_{basepred}(j))^+ \quad (3.8)$$

де h – кількість минулих інтервалів, а x^+ позначає $\max(0, x)$.

3.7 Оцінка часу обслуговування

Двом мережевим функціям з подібними інтенсивностями надходження запитів але різними інтенсивностями обслуговування (наприклад, через різні розміри пакетів і т.д.) потрібно буде виділити різні кількості ресурсів [67].

Для оцінки інтенсивності обслуговування мережевої функції модуль прогнозування обчислює розподіл ймовірностей часу обслуговування запитів. Такий розподіл представляється гістограмою часу обслуговування запиту. По завершенні кожного запиту, ця гістограма оновлюється з урахуванням часу обслуговування цього запиту. Розподіл використовується для визначення очікуваного часу обслуговування s_i для запитів в наступному вікні адаптації. s_i може обчислюватись як середнє, медіана, або процентиль розподілу, отриманого з гістограми. Пропонується використовувати середнє значення розподілу для представлення часу обслуговування запитів мережевої функції.

3.8 Аналіз методу керування ресурсами в контексті зміни інтенсивності надходження запитів

Розглянемо задачу в системі Mathcad. Розглянемо роботу одного блоку і будемо вважати заданою базову компоненту прогнозованої інтенсивності надходження заявок, а також припускаємо наявність ресурсу одного типу.

На рис. 3.6 показано різні варіанти представлення значень часового ряду, тобто навантаження, в залежності від величини базового інтервалу, і їх

вплив на точність представлення навантаження (так помилка репрезентації для випадку інтервалів у 10 хвилин складає 7%, а для випадку 60 хвилин – вже 19%). Для порівняння на рис. 3.7 також зображена система з усього лише 4 вікнами адаптації на день – для неї помилка у представленні навантаження складає 53%, з чого можна зробити ще і той висновок, що ресурси будуть застосовуватись вкрай недоцільно, додатково на рис. 3.8 зображено систему без динамічного керування ресурсами, де помилка дорівнює 235%.

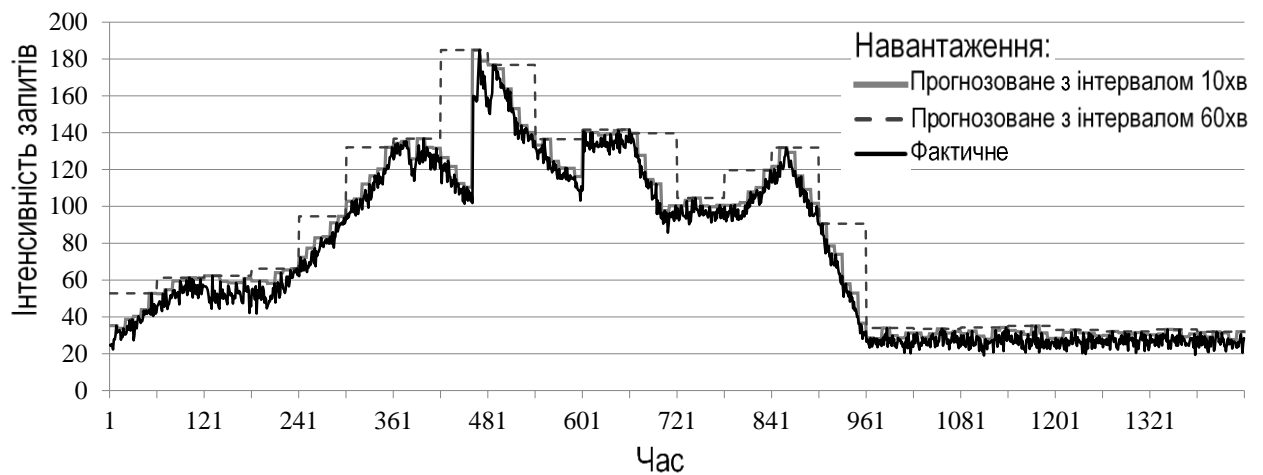


Рисунок 3.6 — Вплив величини інтервалу сталої конфігурації ресурсів на точність представлення очікуваного навантаження

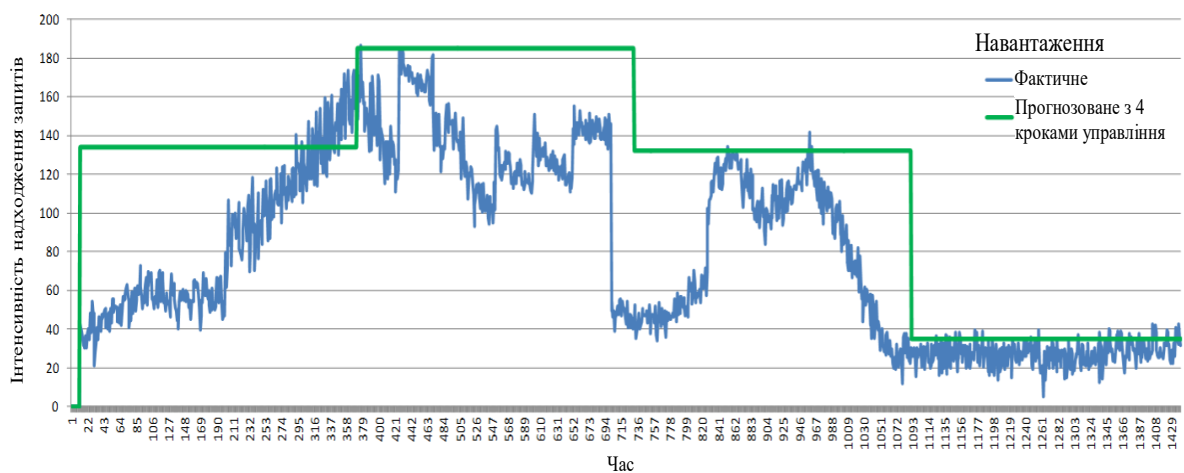


Рисунок 3.7 — Результати моделювання системи з 4 кроками реконфігурації

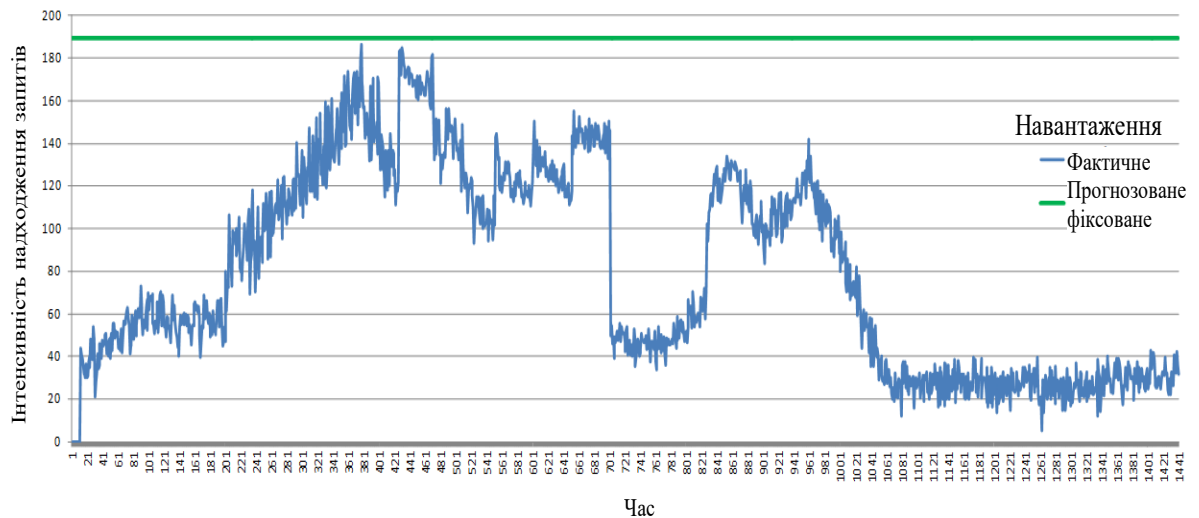


Рисунок 3.8 — Результати моделювання системи з фіксованим керуванням

Результати моделювання системи зі змінною величиною інтервалу часу сталої конфігурації та системи без неї показали (рис. 3.9), що різниця між прогнозованим значенням та репрезентаційним може складати 9%. Якщо не застосовувати систему динамічного регулювання величини вікна керування, то відхилення складатиме 18%, тобто на 9% більше і ресурсів буде витрачатись, відповідно, більше.

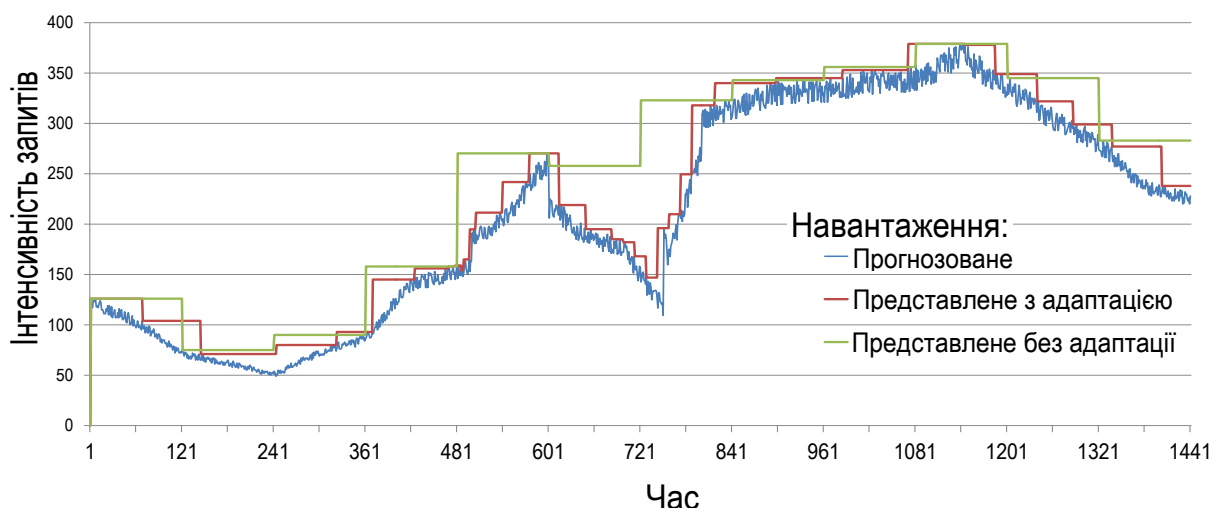


Рисунок 3.9 — Результати моделювання системи з динамічною зміною величини інтервалу часу сталої конфігурації та системи без неї

3.9 Експериментальне дослідження використання ресурсів у віртуалізованій базовій мобільній мережі

Експериментальне дослідження проводиться засобами імітаційного моделювання середовища Matlab. Розглядається система з 5 дата центрів. Для оцінки запропонованого підходу середня кількість вільних ресурсів в день визначалась як різниця між фіксовано виділеними ресурсами, тобто коли протягом доби завжди виділено 100% ресурсів, та динамічно виділеними ресурсами з використанням NFV. Як видно з таблиці 3.1, об'єм ресурсів, що виділяються динамічно, в середньому на 42% менше, ніж у разі використання традиційного підходу розподілу. На рис. 3.10 зображено результат роботи динамічного розподілу ресурсів у віртуалізованому ЕРС мобільної мережі у графічному вигляді. Сіра штрихова лінія ілюструє випадок статичного розподілу ресурсів. Сіра крива показує трафік протягом дня. Чорна крива показує обсяг ресурсів, розподілених відповідно до запропонованого методу динамічно.

Таблиця 3.1 — Середній відсоток вільних ресурсів протягом дня

Дата центр	Середній % вільних ресурсів протягом дня
1	40
2	38
3	42
4	46
5	45
В середньому	42

В роботах [115-118]) здобувачем висвітлено отримані у розділі результати, а саме модель конфігурації ресурсів обслуговування гібридних телекомунікаційних сервісів для забезпечення належного використання ресурсів та якості обслуговування та метод побудови системи керування ресурсами мережевих функцій в центрах обробки даних оператора мережі мобільного зв'язку.

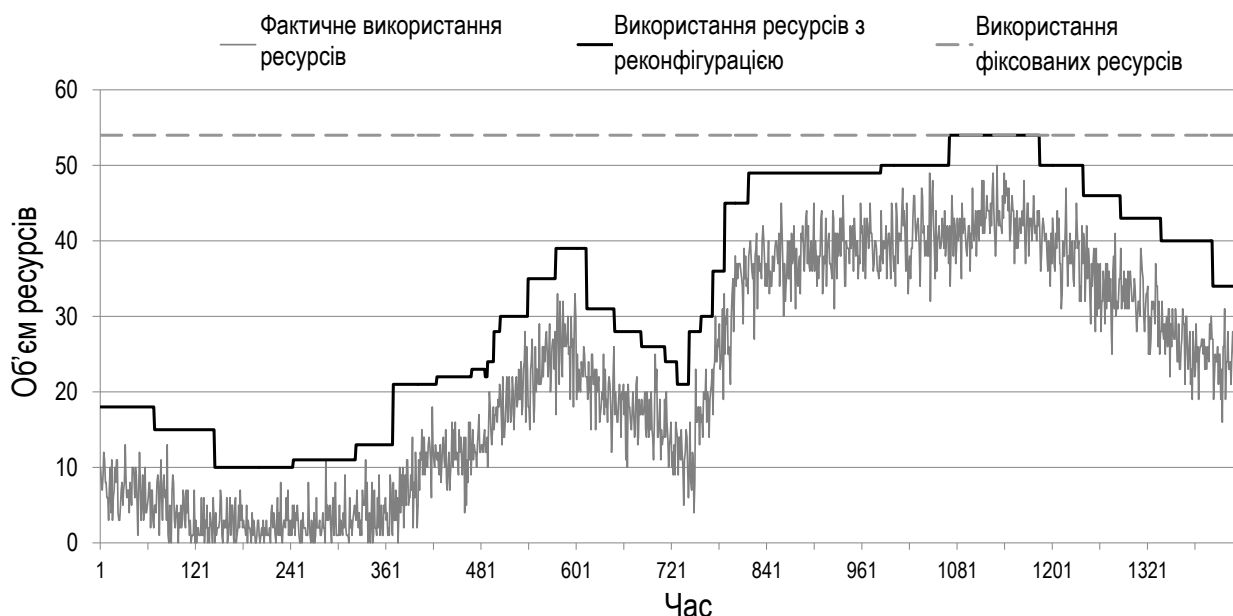


Рисунок 3.10 — Результат динамічного розподілу ресурсів у віртуалізованій базовій мережі

3.10 Автоматизація процесу масштабування мережевих функцій

На рис. 3.11 [45] показано обмін повідомленнями робочого процесу для операції масштабування одного VNF-екземпляра NS. Ці повідомлення були згруповані в різні етапи: збір інформації, запуск масштабування, розподіл ресурсів та звільнення ресурсів. Хоча останні два етапи можуть виконуватися незалежно, може бути, що обидва потрібні в одному сценарію масштабування (наприклад, заміна виконуваного екземпляру VNFC іншим екземпляром більшої ємності). У цьому випадку розподіл відбувається до звільнення, щоб гарантувати безперервність сервісу (наприклад, при запуску нового екземпляру VNFC старий екземпляр можна видалити).

На етапі збору інформації NFVO збирає дані про продуктивність/несправності від VNFMs та VIMs. Показники ефективності на рівні NS/VNF повідомляються за допомогою Performance Information Available Notifications – доступних сповіщень про інформацію продуктивності, та пороговими значеннями, оснований на показниках

продуктивності, за допомогою порогового повідомлення Threshold Crossed Notification (етапи 1-2). Значення індикатора VNF змінюються за допомогою EMs, а також повідомляються NFVO за участі VNFM (етап 3).

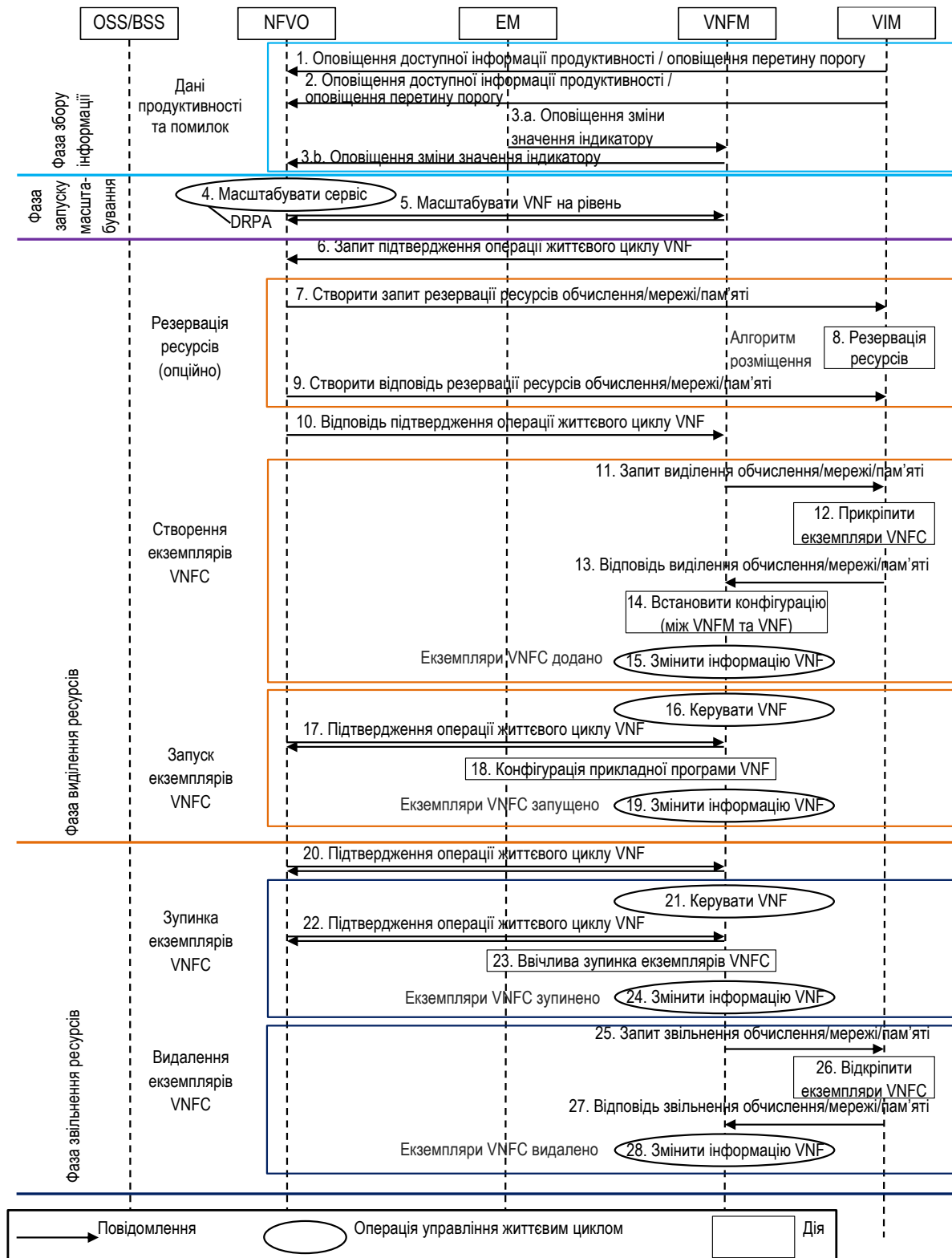


Рисунок 3.11 — Робочий процес процедури масштабування VNF

У фазі запуску масштабування (етап 4) DRPA використовує вказану вище інформацію разом із інформацією, до якої NFVO має доступ зі сховищ даних (NSD/VNFD, NS/VNF Info та стан ресурсів NFVI), щоб визначити оптимальний рівень функціонування (Instantiation Level – IL) NS-IL. Якщо ми вважатимемо, що оптимальний NS-IL відрізняється від існуючого в VNF-IL одного екземпляра VNF, станеться процедура масштабування VNF.

NFVO вимагає, щоб VNFM масштабував екземпляр VNF, надіславши йому новий VNF-IL у повідомленні масштабування Scale VNF to Level Request.

Тепер VNFM може ініціювати операцію масштабування. З цією метою, VNFM забезпечує цю операцію управління життєвим циклом унікальним ідентифікатором, використовуючи повідомлення Scale VNF to Level Response (етап 5). VNFM використовуватиме цей ідентифікатор, щоб повідомити NFVO про початок і пізніше результат цієї операції. Нарешті, VNFM консультується з VNF Info та VNFD для порівняння поточного VNF-IL з новим. З цього порівняння можна визначити ресурси, що підлягають виділенню та/або звільненню для цієї операції масштабування.

На етапі розподілу ресурсів виділяються наступні підфази.

Резервування ресурсів (необов'язково [46]): Перед цим етапом VNFM запитує у NFVO дозвіл на виділення ресурсів. З цією метою VNFM надсилає NFVO ідентифікатори VDU та внутрішніх VL, які вказують на ресурси, які будуть виділятися (етап 6). Хоча NFVO вже мав цю інформацію після кроку 5, специфікації ETSI передбачають обмін такою інформацією [119]. Тоді розпочинається підфаза резервування ресурсу. З результатів DRPA, NFVO знає, які NFVI-PoP повинні мати ресурси, що підлягають розподілу, а також VIM, що забезпечують доступ до цих NFVI-PoP. Тепер ці ресурси можуть бути зарезервовані для подальшого виділення. Кожен вибраний VIM отримує три запити на резервування (етап 7), по одному для кожного типу ресурсу (обчислення, зберігання та мережа), який він повинен зарезервувати в NFVI-PoP під його управлінням. Ці запити містять обмеження щодо місця

розташування, що застосовуються до зазначених ресурсів. VIM використовує ці обмеження для виконання алгоритму розміщення (етап 8), приймаючи рішення щодо відповідних зон ресурсів NFVI-PoP [120], де ресурси зарезервовані. Потім VIM надсилає NFVO (етап 9) ідентифікатори зарезервованих ресурсів. Нарешті, NFVO надсилає VNFM ці ідентифікатори та інформацію про підключення [119] для кожного вибраного VIM. Тепер VNFM знає, як отримати доступ до цих VIM, і який може виділити кожен ресурс. Якщо ця підфаза не виконується, необхідно розглянути два питання. По-перше, виконуються лише кроки 6 і 10. По-друге, алгоритм розміщення тепер виконується після запиту розподілення ресурсів (див. наступну підфазу).

Створення екземплярів VNFC: VNFM надсилає ідентифікатори резервування відповідним VIM (етап 11) для виділення ресурсів (крок 12). На цьому етапі створено екземпляри VNFC та підключено їх. Ідентифікатори виділених ресурсів потім надсилаються до VNFM (етап 13). На кроці 14, VNFM запускає конфігурацію екземплярів VNFC. Нарешті, VNFM оновлює інформацію VNF у сховищі даних (етап 15), щоб відобразити створення нових екземплярів VNFC та встановлює їх стан на STOPPED.

Запуск екземплярів VNFC: Щоб запустити функціональність нових екземплярів, VNFM запускає операцію життєвого циклу Operate VNF (етап 16). Ця операція змусить (наприкінці цієї підфази) змінити стан екземплярів з STOPPED на STARTED. Коли NFVO підтверджує цю операцію (етап 17), нові екземпляри VNFC налаштовуються на рівні програми (етап 18). З цією метою VNFM спілкується з ЕМ. Нарешті, VNFM оновлює інформацію VNF (етап 19), змінюючи стан нових екземплярів з STOPPED на STARTED. Слід зазначити, що деякі з виконуваних екземплярів VNFC можуть зазнати впливу через створення нових, і, отже, вони повинні бути (ре)конфігуровані з точки зору підключення (наприклад, нові інтерфейси, оновлені вимоги до каналу) та/або програми (наприклад, надсилення/отримання пакетів до/з нових екземплярів). У такому випадку

VNFM буде вимагати від відповідних VIM та/або ЕМ, відповідно, внесення необхідних змін. Для фази Вивільнення ресурсу виділяють дві підфази.

Зупинка екземплярів VNFC: На кроці 20 VNFM просить NFVO дозволу на звільнення ресурсів. Тоді VNFM запускає операцію життєвого циклу Operate VNF (етап 21), щоб чомно зупинити деякі екземпляри VNFC (змушуючи зупинитись екземпляр VNFC в кінці цієї підфази). На кроці 23, затронуті екземпляри (ре)конфігуруються (слідуючи відповідно до стратегій, що зазначені у підфазі Запуску екземплярів VNFC), а екземпляри, що підлягають завершенню, вимикаються. Нарешті, стан зупинених екземплярів змінюється з STARTED на STOPPED (етап 24).

Видалення екземплярів VNFC: VNFM надсилає відповідним VIM ідентифікатори ресурсів, які розміщують та з'єднують зупинені екземпляри VNFC (етап 25). На даному етапі ці екземпляри видаляються (етап 26). Потім VIM надсилають VNFM ідентифікатори звільнених ресурсів (етап 27). Після отримання цих ідентифікаторів VNFM оновлює інформацію VNF (етап 28), щоб відобразити видалення екземпляра.

Висновки

1. Коректне керування обчислювальними ресурсами в дата центрах для задоволення вимог до якості обслуговування при цьому мінімізуючи споживання ресурсів є складним завданням.

2. У присутності динамічно змінюваних навантажень потрібні методи динамічного керування конфігурацією ресурсів, для того щоб забезпечити гарантії для віртуалізованих мережесих функцій, що працюють на загальних дата центрах. Для вирішення цієї проблеми використовуємо системну архітектуру, яка поєднує в собі онлайн вимірювання з методами прогнозування і розподілу ресурсів.

3. Розроблено метод визначення розміру інтервалу часу сталої конфігурації ресурсів для точного визначення шаблонів навантаження, який

передбачає змінну величину інтервалів з метою оптимального використання ресурсів та забезпечення заданої якості обслуговування.

4. Запропоновано метод, який поєднує прогностичне і реактивне надання ресурсів: прогностичне надання ресурсів виділяє ресурси завчасно в очікуванні певного пікового навантаження, в той час як реактивне надання ресурсів вносить корективи після того, як спостерігалось аномальне збільшення навантаження. Іншими словами, у той час як прогностичне надання ресурсів намагається «випереджати» коливання навантаження, реактивне надання ресурсів дозволяє хостинговій платформі бути гнучкою до відхилень від очікуваного навантаження.

РОЗДІЛ 4

РЕКОНФІГУРАЦІЯ МЕРЕЖІ ПІСЛЯ ПЕРЕВАНТАЖЕННЯ АБО ЗБОЮ

4.1 Модель мережі та постановка задачі

Фізична мережа задана у вигляді графа $SN=(N,NE)$, де N є множиною фізичних вузлів і NE – множиною каналів. Кожен канал $(n_1,n_2) \in NE$, $n_1,n_2 \in N$ має максимальну пропускну здатність $c(n_1,n_2)$ і мережеву затримку $L(n_1,n_2)$, а кожен вузол $n \in N$ пов'язаний з певними ресурсами c_n^i , $i \in R$, де R – множина типів ресурсів. Мережа зв'язку представлена множиною сервісів (або запитів віртуальної мережі) K , які вбудовуються в фізичну мережу. Запит віртуальної мережі k , $k \in K$, можна представити як зважений граф $G_k=(V_k,E_k)$, де V_k є множиною віртуальних вузлів, що містить h_k елементів і позначається як $V_k=(v_{k,1},v_{k,2},\dots,v_{k,h_k})$, де $v_{k,j}$ означає j -у мережеву функцію у ланцюзі функцій k . E_k є множиною віртуальних каналів $e_k(v_{k,j},v_{k,g}) \in E_k$. Вимоги смуги пропускання каналу між двома функціями, $j1$ і $j2$, що відносяться до сервісу $k \in K$ позначається як $d_k^{(j1,j2)}$, $d_k^{j,i}$ – кількість ресурсу типу i , що виділяється для мережевої функції j сервісу k . Булеві змінні $x_n^{k,j}$ вказують, чи мережева функція j , пов'язана з $k \in K$, розташовується на фізичному вузлі n , змінні $f_{(n_1,n_2)}^{k,(j1,j2)}$ визначають, чи фізичний канал (n_1,n_2) використовується у шляху між $j1$ та $j2$ для запиту k . L_k – максимальна затримка для запиту k . $costN(i,n)$ – вартість зайнятої одиниці ресурсу i на фізичному вузлі n , і $costL(n_1,n_2)$ – вартість зайнятої одиниці пропускну здатності на фізичному каналі $(n_1,n_2) \in NE$. $suit_n^{k,j}$ означає що функція j з запиту k може бути розміщена на вузлі n .

MN являє собою множину вузлів керування, де $MN \subseteq N$, які відповідають за функціонування пропонованого механізму реконфігурації після відмови або перевантаження. Кожен керуючий вузол пов'язаний з одним або декількома вузлами фізичної мережі і виконує кроки, необхідні для

відновлення належного функціонування мережі. Процес призначення вузлів керування і критерії, які враховуються при виборі вузлів керування, будуть досліджені далі.

У запропонованому підході припускаємо, що відображення запитів віртуальної мережі вже виконано і вузли керування у фізичній мережі призначаються, як описано нижче.

Процес відображення віртуальної мережі відбувається в два етапи [87]: відображення вузлів ($M_N : V_k \rightarrow N$) і відображення каналів ($M_L : E_k \rightarrow NE$).

Проблема, що розглядається далі, полягає в тому, як перемістити розміщені на вузлі з відмовою віртуальні вузли з метою мінімізації витрат відновлення вузла відмови і періоду переривання сервісу. У запропонованому підході підкреслюється, що будь-яке переміщення віртуального вузла повинно виконуватись локально і бути координованим тільки вузлами керування.

4.2 Оптимальне розміщення вузлів керування у мережах, заснованих на NFV

Припускаємо, що вузли керування (далі – менеджери) можуть розміщуватись в вузлах N .

При заданій кількості менеджерів A існує скінченна множина з $\binom{|N|}{A}$ можливих розташувань, відповідно, задача розміщення менеджерів є задачею багатокритеріальної комбінаторної оптимізації. Метою задачі є знаходження таких розташувань менеджерів з множини можливих розташувань розміру A : $Pl_A = \{Pl \in 2^{|N|} \mid |Pl| = A\}$, що є оптимальними відповідно до деякої цільової функції.

Метою оптимізації є визначення місця розташування кожного менеджера при заданій їх кількості A , так що мінімізується функція загальних витрат $Cost_A(\{p_n : n \in N\})$, де p_n – булева змінна, яка рівна одиниці, якщо менеджер розміщується в точці n . Задача оптимізації буде мати вигляд (4.1).

$$\min_{\{p_n: n \in N\}} U_A$$

при обмеженні $\sum_n p_n = A$

(4.1)

Основною метою хорошого розміщення менеджерів є мінімізація затримок між вузлами і менеджерами в мережі. Проте, розглядати тільки затримки не достатньо. Розміщення менеджерів повинно також враховувати певні обмеження стійкості.

На рис. 4.1 показані різні питання, які необхідно враховувати при оцінці стійкості розміщення. Нижче коротко пояснимо ці питання, і що необхідно, щоб бути стійким по відношенню до них. На рис. 4.1 показано нормалізовані затримки між вузлами та інтенсивність навантаження на вузлах.

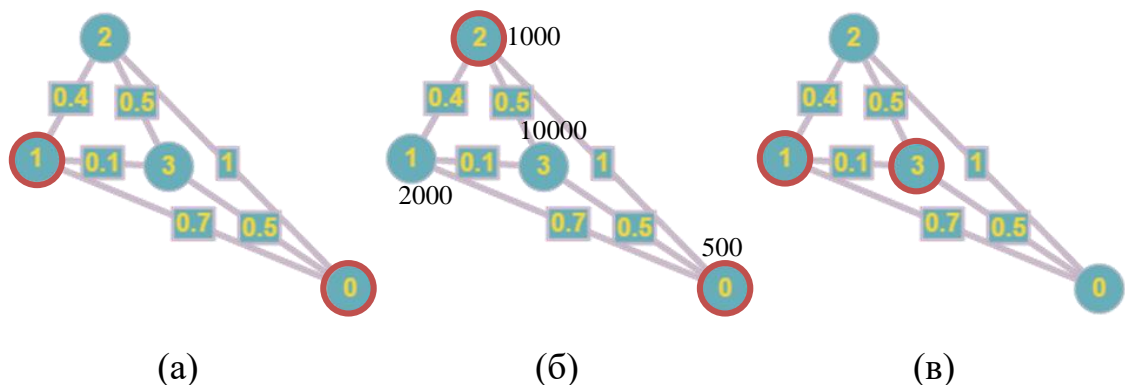


Рисунок 4.1 — Розміщення по менеджерів за різним критерієм: (а) мінімальної затримки до менеджера; (б) мінімального дисбалансу навантаження на менеджерів; (в) мінімальної затримки між менеджерами

Аналогічно до [65], припустимо, що вузли призначаються до їх найближчого менеджера, використовуючи в якості метрики затримки $dl_{g1,g2}$ між вузлом $g1$ і менеджером $g2$. Кількість вузлів на менеджера може бути незбалансованою – чим більше вузлів менеджер повинен контролювати, тим вище навантаження на цього менеджера. Якщо кількість запитів вузла до менеджера в мережі збільшується, аналогічно поводить себе і ймовірність додаткових затримок через черги в системі керування. Для того, щоб бути

стійким від перевантаження менеджера, призначення вузлів різним менеджерам повинно бути добре збалансованим.

Цілком очевидно, що одного менеджера не досить, щоб досягти стійкості в мережі. Проте, коли кілька менеджерів розміщуються в мережі, логіка керування мережі розподіляється по декількох менеджерах і ці менеджери повинні синхронізуватися, щоб підтримувати несуперечний глобальний стан. Залежно від частоти синхронізації між менеджерами, затримка між окремими менеджерами грає важливу роль.

4.2.1 Затримка вузол-менеджер

Аналогічно, як представлено в [64], на основі матриці dl , що містить відстані найкоротших шляхів між усіма вузлами, максимальний час затримки передачі між вузлом і менеджером для певного розміщення менеджерів може бути визначений як

$$U_A^{latency}(p) = \max ddc_n, \quad (4.2)$$

де ddc_n – максимальна затримка передачі від вузла мережі до менеджера в точці n ; ddc_n розраховується наступним чином:

$$ddc_n = \max_{g \in N} latency_g \cdot \pi_{g,n},$$

де $latency_g$ – затримка між менеджером та вузлом g , $latency_g = \min_{\{n: n \in N \cap p_n = 1\}} dl_{g,n}$;

$\pi_{g,n}$ – булева змінна, яка рівна одиниці, якщо вузол g обслуговується менеджером розміщеним в точці n .

Розглядаємо не середнє, але максимальне значення затримки, так як середнє приховує значення найгіршого випадку, які є важливими, коли стійкість повинна бути поліпшена.

4.2.2 Збалансований розподіл навантаження менеджерів

Залежно від ситуації, може бути бажано мати приблизно рівне навантаження на всіх менеджерів, так що жоден менеджер не перевантажується, а у інших мало роботи. Далі розглядаємо збалансований розподіл вузлів між менеджерами. Як формальну метрику вводимо баланс розміщення або, вірніше, дисбаланс, $U_A^{imbalance}$, тобто відхилення від повністю збалансованого розподілу, як різниця між навантаженням на найбільш завантаженому менеджері і найменш завантаженому менеджері.

$U_A^{imbalance}$ визначається в такий спосіб:

$$U_A^{imbalance}(p) = \max ldc_n - \min ldc_n, \text{ де } ldc_n > 0, \quad (4.3)$$

ldc_n – навантаження на менеджер в точці n ; ldc_n розраховується наступним чином:

$$ldc_n = \sum_{g \in N} load_g \cdot \pi_{g,n},$$

де $load_g$ – коефіцієнт навантаження на вузол g .

4.2.3 Затримка менеджер-менеджер

Як останній аспект стійкого розміщення менеджерів, розглянемо як затримка між менеджерами може враховуватися при виборі розміщення менеджерів. Формально, затримка між менеджерами $U_A^{interlatency}$ визначається як найбільша затримка між будь-якими двома менеджерами при заданому розміщенні:

$$U_A^{interlatency}(p) = \max_{\{n,g:n,g \in N \cap p_n=1, p_g=1\}} dl_{g,n}. \quad (4.4)$$

Загалом, розміщення з урахуванням затримки між менеджерами мають тенденцію розміщувати всіх менеджерів набагато ближче один до одного. Це збільшує максимальну затримку від вузлів до менеджерів.

4.2.4 Цільова функція оптимізації

Таким чином,

$$U_A = wu^{latency} \times U_A^{latency}(p) + wu^{imbalane} \times U_A^{imbalane}(p) + wu^{interlatency} \times U_A^{interlatency}(p), \quad (4.5)$$

де wu – множина вагових коефіцієнтів.

На рис. 4.2 показані можливі рішення по розміщенню двох менеджерів у мережі з 10 вузлами. Оптимальне значення показує, що усі оптимізаційні цілі не можуть бути досугнуті одночасно.

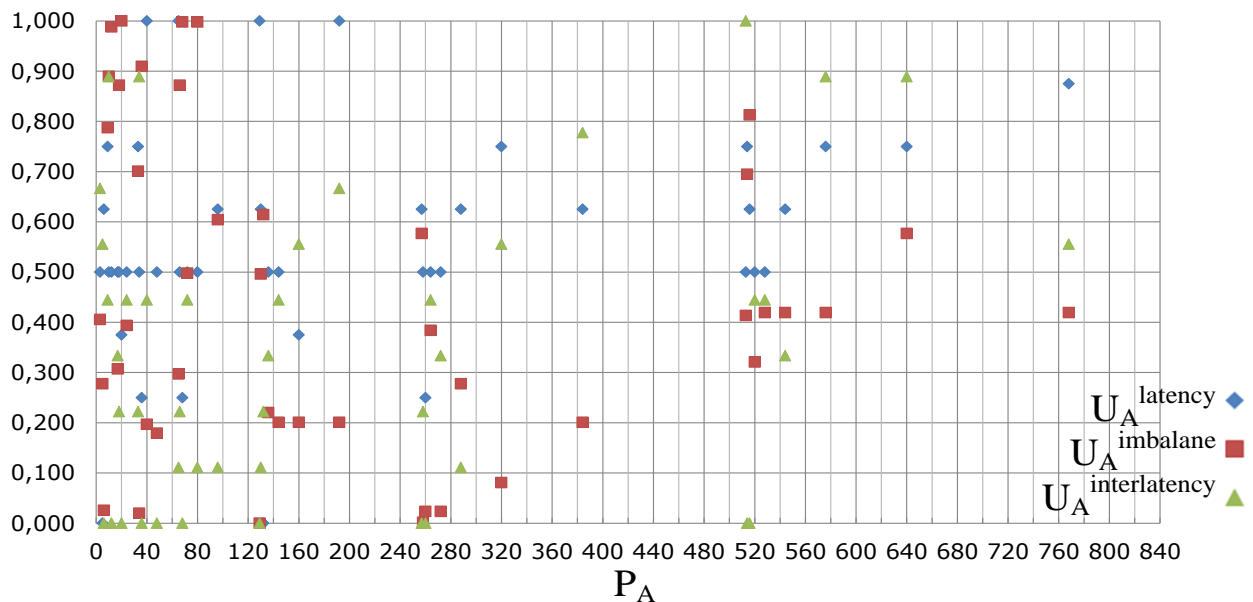


Рисунок 4.2 — Простір рішень оптимізаційної задачі розміщення менеджерів

4.2.5 Конфігурація вузлів менеджерів

Кожен менеджер (наприклад, вузол 3 на рис. 4.3 управляє вузлом 2) підтримує таблицю (наприклад, таблиця 4.1), що містить ідентифікатори віртуальних мереж, які мають віртуальні вузли розміщені на керованому фізичному вузлі [90]. Крім того, таблиця містить посилання на розміщені на

фізичних вузлах віртуальні вузли (наприклад, вузли c та d на рис. 4.3) і суміжні до них віртуальні вузли (наприклад, вузли a та b на рис. 4.3). Ця таблиця безперервно оновлюється (наприклад, кожен раз, коли приймається або відхиляється запит віртуальної мережі).

Таблиця 4.1 — Таблиця на менеджері 3 на рис. 4.3

ID віртуальної мережі	Віртуальний вузол	Ємність	Прилеглі вузли
1	c	10	a на вузлі 0 b на вузлі 1
2	d	15	f на вузлі 4 e на вузлі 5

На рис. 4.3 показаний приклад фізичної мережі з двома вже вбудованими віртуальними мережами. Вузол з відмовою (вузол 2) був призначений менеджеру (вузлу 3), який відповідає за переміщення віртуальних вузлів.

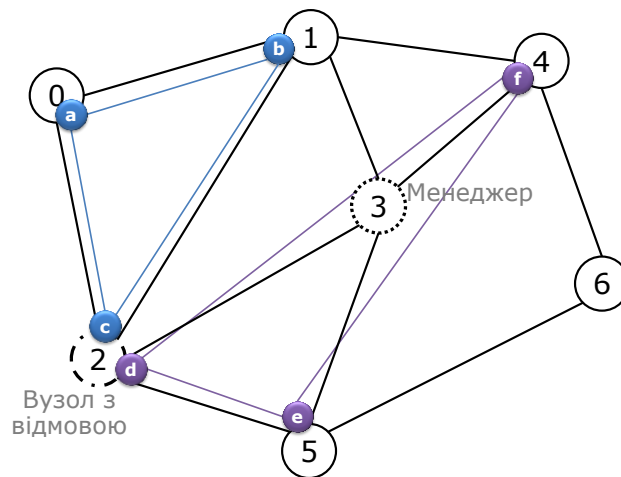


Рисунок 4.3 — Фізична мережа з вузлом з відмовою, на якому розміщується два віртуальних вузла

4.3 Схема відновлення мережі після відмови

Процес переміщення вузлів віртуальної мережі, розміщених на вузлі, який відмовив, $v_{k,f}^{fail}$, запускається, коли система відправляє запит на відновлення відповідному вузлу-менеджеру (наприклад, вузлу 3 на рис. 4.3).

Процес відновлення для кожного порушеного вузла віртуальної мережі протікає в такий спосіб: менеджер направляє запит на відновлення до всіх вузлів фізичної мережі, на яких розміщуються віртуальні вузли, суміжні з ураженими віртуальними вузлами (наприклад, вузол 0 і вузол 1 на рис. 4.3). Кожен з цих вузлів будує дерево найкоротших шляхів (Shortest Path Tree – SPT) до всіх вузлів фізичної мережі на відстані не більше l (поріг встановлюється провайдером послуг), де коренем SPT виступає сам цей вузол. Менеджер використовує ці шляхи, щоб вибрати вузол з оптимальною відстанню до всіх вузлів фізичної мережі, де розташовані вузли віртуальної мережі прилеглі до несправного вузла. Цей вузол в кінцевому рахунку стає оптимальним кандидатом для розміщення віртуального вузла з відмовою. Крім того, ємність кінцевих вузлів шляхів з SPT повинна бути не менше ємності віртуального вузла, розміщеного на вузлі з відмовою (наприклад, вузла с). Обираємо вузол з мінімальною вартістю шляху до всіх кореневих вузлів у деревах SPT та мінімальною вартістю обчислень. Рис. 4.4 містить опис псевдокоду алгоритму відновлення вузла після відмови і виконується для всіх $\{v_{k,j} : x_n^{k,j}=1 \ \& \ n = failed\}$.

4.4 Схеми відновлення у разі перевантаження вузла

У мережі існує ймовірність відмови вузла через перевантаження. Для виконання відновлення при відмові вузла в перевантаженій мережі, виконується процедура реконфігурації для міграції віртуальних вузлів, розміщених на перевантаженому фізичному вузлі (рис. 4.5).

Процес відновлення починається з сортування всіх віртуальних вузлів, розміщених на перевантаженому фізичному вузлі. Критерієм (CRT в Алгоритмі рис. 4.6), що використовуються для сортування цих вузлів віртуальної мережі, є ємність віртуальних вузлів. Потім виконується процедура відновлення на першому відсортованому вузлі віртуальної мережі,

що має ємність рівну перевантаженню, для переміщення на новий вузол фізичної мережі.

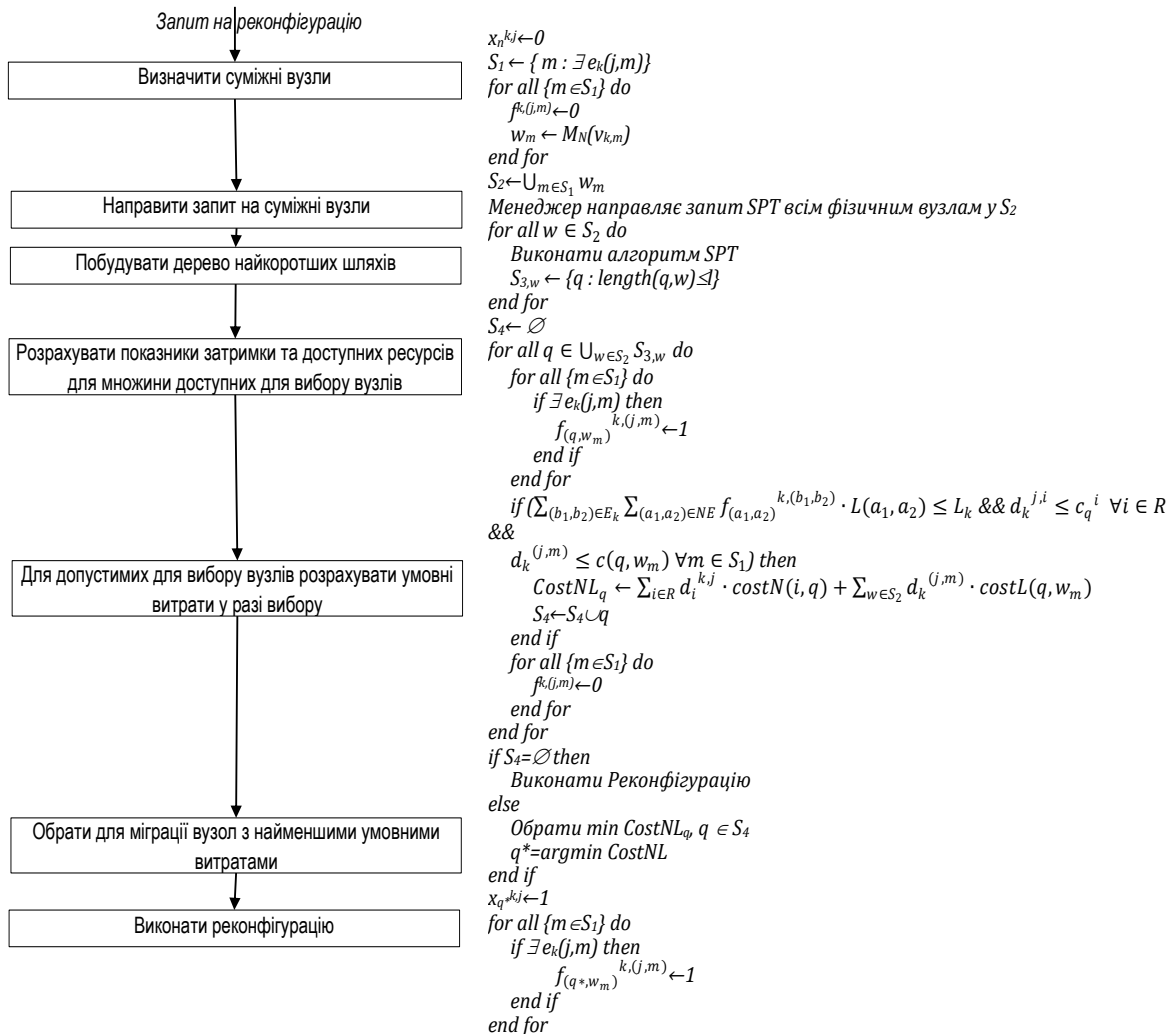


Рисунок 4.4 — Алгоритм відновлення вузла з відмовою



Рисунок 4.5 — Схема відновлення вузла з перевантаженням

$n = \text{перевантажений}$

$S_1 \leftarrow$ Відсортувати віртуальні вузли, що розміщуються на n у зростаючому порядку на основі критерію CRT

Вибрати з S_1 перший віртуальний вузол $v_{k,j}$, ємність ресурсів не менше перевантаженої ємності $d_k^{j,i} \geq \Delta c_n^i \ \forall i \in R$

Виконати алгоритм Відновлення Вузла

Рисунок 4.6 — Алгоритм відновлення вузла з перевантаженням

4.5 Реконфігурація перевантажених мереж

Коли навантаження або ресурси змінюються, деякі VNF можливо доведеться перемістити. Існує ймовірність того, що знайти новий вузол-кандидат для вузла віртуальної мережі, розміщеного на вузлі з відмовою, не вийде. В такому випадку виконується процедура реконфігурації для міграції одного або декількох віртуальних вузлів для переміщення розміщених вузлів віртуальної мережі. Розглянемо задачу міграції як задачу оптимізації, яка спрямована на мінімізацію загальної вартості міграції при обмеженнях допустимої затримки і обчислювальних ресурсів.

Метою оптимізації є знаходження розташування ланцюгів сервісів мережі (тобто розміщення мережевих функцій та розподіл ресурсів), так щоб мінімізувати витрати на зайняті ресурси каналів і вузлів у фізичній мережі, при цьому задовільняючи вимоги трафіку. Сформулюємо цільову функцію у вигляді лінійної комбінації (з ваговими коефіцієнтами a , b , c , e) вартісних виразів.

Визначимо бінарну змінну $x_n^{k,j} \in \{0,1\}$, для позначення того, що VNF j пов'язана з ланцюгом сервісу k розміщується на вузлі n після міграції. Індикатор $x_n^{k,j}=0$ означає, що VNF j не розміщується на вузлі n після міграції; в іншому випадку j розміщується на вузлі n після міграції.

Введемо бінарну змінну $y_n^{k,j}$ для індикації стану мережі перед міграцією. Вона схожа зі змінною $x_n^{k,j}$, $y_n^{k,j}=0$ означає, що VNF j сервісу k не перебуває на вузлі n до міграції; в іншому випадку, j розташована на вузлі n до міграції.

Таким чином, можемо використовувати індикатор $I^{k,j}$, щоб вказати чи VNF j сервісу k було переміщено в поточному процесі міграції.

$$I^{k,j} = \sum_{n \in N} x_n^{k,j} \cdot y_n^{k,j}$$

$I^{k,j} = 0$ вказує, що VNF було переміщено в поточному процесі міграції, і $I^{k,j} = 1$ вказує, що VNF не було переміщено.

x_n позначає чи n -ий фізичний сервер працює або ні після міграції.

$$x_n = \begin{cases} 1 & (\text{сервер } n \text{ працює}) \\ 0 & (\text{в іншому випадку}) \end{cases}$$

y_n позначає чи n -ий фізичний сервер працює або ні перед міграцією.

$$y_n = \begin{cases} 1 & (\text{сервер } n \text{ працює}) \\ 0 & (\text{в іншому випадку}) \end{cases}$$

Для того щоб розглянути ресурси, які споживаються при міграції та запуску серверів, вводимо такі вирази:

– B_n позначає необхідні витрати b_n для запуску n -ого серверу.

$$B_n = b_n x_n (x_n - y_n);$$

– $L_i^{k,j}(n \rightarrow n')$ позначає використання ресурсу i для міграції VNF j з ланцюгу сервісу k з серверу n на сервер n' .

$$L_i^{k,j}(n \rightarrow n') = l_i(d^{k,j}) + l'_i(d^{k,j}),$$

де $l_i(x)$ – функція використання ресурсу i для міграції з серверу,

$l'_i(x)$ – функція використання ресурсу i для міграції на сервер.

Цільова функція буде визначатися як:

$$\begin{aligned} MCost = & a \cdot \sum_{n \in N} (B_n + x_n \cdot cost(n)) + \\ & + b \cdot \sum_{n \in N} \sum_{k \in K} \sum_{j \in V} \sum_{i \in R} x_n^{k,j} \cdot d_i^{k,j} \cdot costN(i, n) + \\ & + c \cdot \sum_{(n_1, n_2) \in NE} costL(n_1, n_2) \cdot \sum_{k \in K} \sum_{(j_1, j_2) \in E} f_{(n_1, n_2)}^{k, (j_1, j_2)} \cdot d_k^{(j_1, j_2)} + \\ & + e \cdot \sum_{n \in N} \sum_{n' \in N} L_i^{k,j}(n \rightarrow n') x_{n'} (x_{n'} - y_n) \end{aligned} \quad (4.6)$$

Використовуючи наведені вище міркування, формулюємо задачу наступним чином.

Цільова функція:

$$\text{Min } M\text{Cost}$$

З обмеженнями:

$$\sum_{n \in N} x_n^{k,j} = 1 \quad \forall k \in K, j \in V, \quad (4.7)$$

$$x_n^{k,j} \leq \text{suit}_n^{k,j} \quad \forall k \in K, j \in V, n \in N, \quad (4.8)$$

$$\sum_{k \in K} \sum_{j \in V} x_n^{k,j} \cdot d_i^{k,j} + y_n^{k,j} \cdot (1 - I^{k,j}) \cdot l_i(d^{k,j}) + x_n^{k,j} \cdot (1 - I^{k,j}) \cdot l'_i(d^{k,j}) \leq c_n^i \quad \forall n \in N, i \in R, \quad (4.9)$$

$$\sum_{t \in K} \sum_{(j_1, j_2) \in E} f_{(n_1, n_2)}^{k, (j_1, j_2)} \cdot d_k^{(j_1, j_2)} \leq c(n_1, n_2) \quad \forall (n_1, n_2) \in NE, \quad (4.10)$$

$$\sum_{(n, w) \in L} f_{(w, n)}^{k, (j_1, j_2)} - f_{(n, w)}^{k, (j_1, j_2)} = x_n^{k, j_1} - x_n^{k, j_2} \quad \forall k \in K, n \in N, (j_1, j_2) \in E, \quad (4.11)$$

$$x_n^{k,j}, f_{(n_1, n_2)}^{k, (j_1, j_2)} \in \{0, 1\} \quad \forall k \in K, j \in V, n \in N, (j_1, j_2) \in E, (n_1, n_2) \in NE, \quad (4.12)$$

$$\sum_{(j_1, j_2) \in E} \sum_{(n_1, n_2) \in NE} f_{(n_1, n_2)}^{k, (j_1, j_2)} \cdot L(n_1, n_2) \leq L_k \quad \forall k \in K, \quad (4.13)$$

$$\sum_{n \in N} x_n^{k,j} \sum_{i \in R} \left(\frac{1}{\frac{d_k^{j,i}}{s_{n,i}^{k,j}} - \lambda^{k,j}} \right) \leq P_k^j \quad \forall k \in K, j \in V \quad (4.14)$$

Отже, цільова функція (4.6) представляє собою зважену суму чотирьох вартісних виразів та направлена на мінімізацію: вартості запуску та використання серверу, використання ресурсів серверу, використання каналів зв'язку та використання ресурсів для міграції. Обмеження (4.7) гарантує однократність розміщення мережевих функцій, а (4.8) – адміністративну можливість розміщення на вузлі. Вирази (4.9) та (4.10) являють собою обмеження на ресурси фізичних вузлів і каналів, тобто забезпечують той факт, що кількість задіяних на вузлі ресурсів не перевищує кількості доступних. Вираз (4.11) представляє собою обмеження щодо збереження

поток для всіх шляхів у фізичній мережі, тобто що вхідний потік на вузлі дорівнює вихідному потоку. Вираз (4.12) гарантує, що змінні у задачі є булевими. Вирази (4.13) та (4.14) представляють собою обмеження на час передачі телекомунікаційними каналами та час обробки вузлами обслуговування відповідно, і забезпечують дотримання заданих часових вимог до обслуговування сервісу.

4.6 Оцінка методу локальної реконфігурації мережі

За результатами моделювання (рис. 4.7) запропонований метод показав до 27% менші умовні витрати у порівнянні зі стратегією направленою на мінімізацію затримки, при цьому затримка знаходилась у допустимих межах але була на 20% більшою.

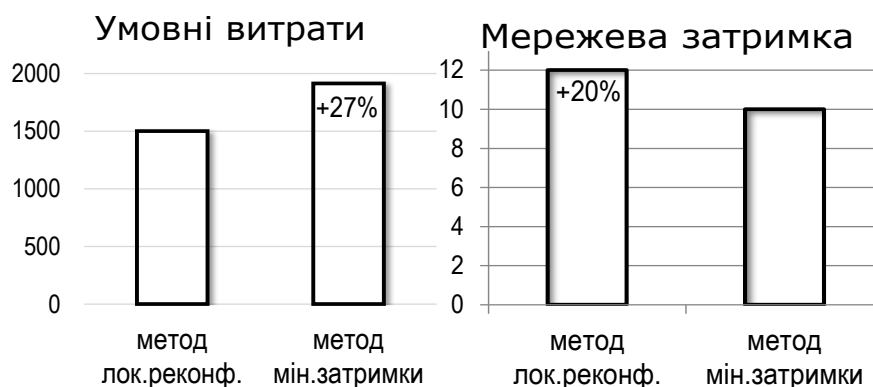


Рисунок 4.7 — Результати відновлення вузла

В роботах [121-122]) описано удосконалені математичні методи з розділу, які використовуються при реконфігурації обчислювальних ресурсів мережі у випадку відмови або перевантаження з використанням технології віртуалізації.

4.7 Модифікована система керування ресурсами

Таким чином, перед початком роботи необхідно мати статистичні дані інтенсивності надходження заявок на мережеву функцію та ймовірнісні характеристики обслуговування заявки. Згідно з методом розміщення

визначається прив'язка кожної мережевої функції традиційної мережі до дата центру та кількості ресурсів, що потрібно зарезервувати для відповідної віртуалізованої мережевої функції. Далі необхідно розбити цикл роботи мережевої функції на інтервали протягом яких її конфігурація буде незмінною і відповідно буде активізовано визначений обсяг ресурсів згідно з методом визначення розміру інтервалу часу сталої конфігурації ресурсів з урахуванням очікуваного навантаження. Під час функціонування мережі мобільного зв'язку може виникнути ситуація, коли фізичний вузол не в змозі надалі обслуговувати вхідне навантаження через недостатність ресурсів для цього або через вихід його з ладу, а такому випадку запускається розподілений метод локальної реконфігурації ресурсів, який перерозподіляє віртуальні вузли у мережі, що постраждали від відмови або перевантаження на фізичному вузлі.

Загальна система керування ресурсами зображена на рис. 4.8.

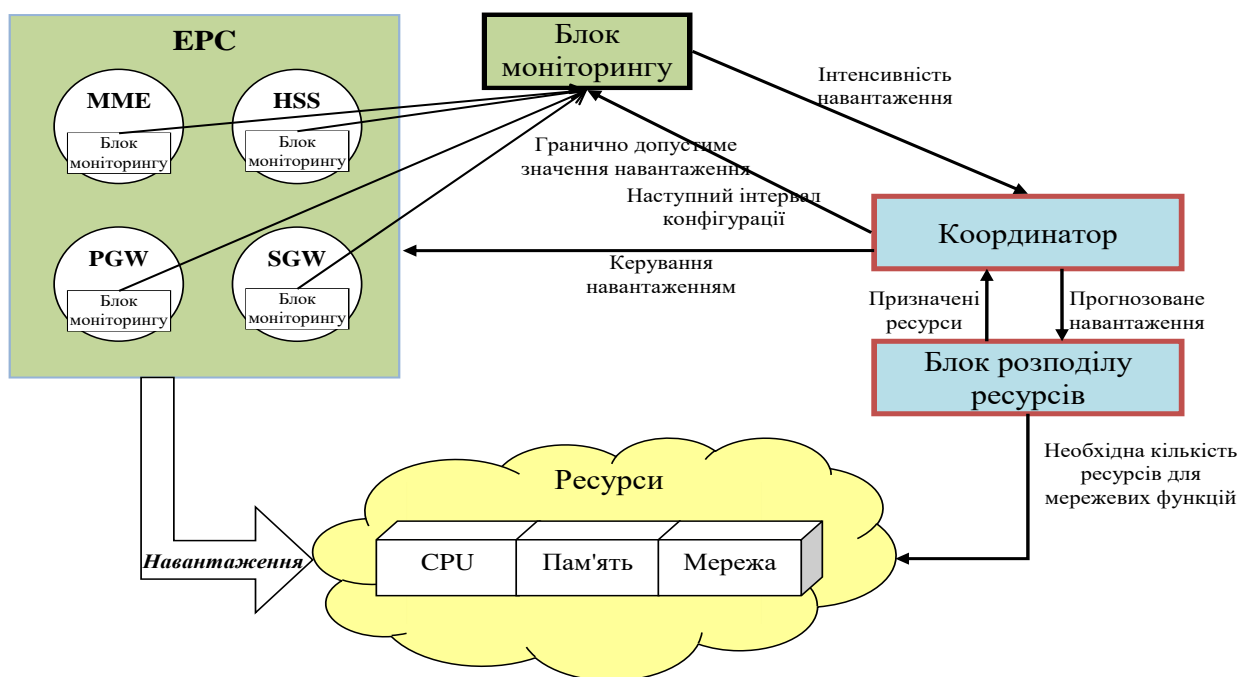


Рисунок 4.8 – Модифікована система керування ресурсами

Система моніторингу відслідковує трафік і підраховує кількість запитів. В системі моніторингу встановлюється порогове значення для числа запитів і до координатора посиляється повідомлення, якщо система моніторингу

виявляє перевантаження. Коли координатор приймає повідомлення, що стосується перевантаження, з системи моніторингу, блок розподілу ресурсів обчислює необхідну кількість ресурсів для обробки заявок належним чином і динамічно розподіляє розрахунковий об'єм. Потім координатор перенаправляє заявки і перевантаження усувається.

Координатор запускається періодично. Для прогнозу базового навантаження, можемо взяти середнє значення історичного денного навантаження. Координатор направляє вхідне навантаження на дата центр, виділений для обслуговування надлишкового службового навантаження, а також обмінюється даними з блоком розподілу ресурсів для надання інформації про передбачене вхідне навантаження.

Блок розподілу ресурсів відповідає за розподіл належної кількості ресурсів, необхідної для обробки навантаження з заданими показниками якості. Протягом безпосереднього функціонування системи цей блок запускається коли реальне навантаження перевищує базове прогнозоване значення навантаження, для того щоб надати додаткові ресурси для надлишкового навантаження. Оскільки блок розподілу ресурсів та координатор не запускаються коли реальне навантаження нижче ніж прогнозоване, процедура реконфігурації ресурсів створює мінімальні додаткові втрати пов'язані з цим процесом.

Загальна схема роботи системи керування ресурсами зображена на рис. 4.9.

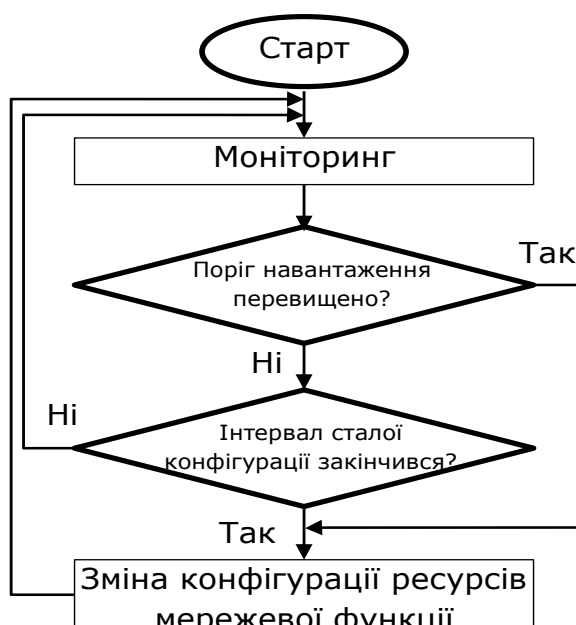


Рисунок 4.9 – Метод керування ресурсами системи

Висновки

1. Розглянуто метод реконфігурації системи у випадку відмови вузла через перевантаження або збій, який спирається на співпрацю множини розподілених керуючих вузлів – менеджерів, розміщених на ряді фізичних вузлів. У дослідженні вирішується проблема розташування локальних вузлів керування з урахуванням параметрів затримки та стійкості, пропонується цільова функція комбінаторної оптимізаційної задачі розташування вузлів керування.

2. Пропонована схема відновлення мережі зменшує вартість відновлення вузла після збоїв, зазнавану затримку і час переривання сервісу під час цього процесу, максимізуючи дохід телекомунікаційного провайдера.

ЗАГАЛЬНІ ВИСНОВКИ

У дисертаційній роботі вирішено важливе наукове завдання пов'язане з організацією обчислювальних ресурсів вузлів базової мобільної мережі та координації потоків між вузлами обслуговування у гібридному середовищі, яке складається із телекомунікаційної мережі та віртуальних обчислювальних сутностей розміщених у хмарі.

В дисертаційній роботі отримані такі теоретичні результати:

1. Проаналізовано структуру телекомунікаційних сервісів, які надаються компаніями мобільного зв'язку, досліджено систему обслуговування службових потоків та підсистем базової мобільної мережі, виявлено, що збільшення службового навантаження призводить до збільшення необхідності у додаткових ресурсах, разом з тим нерівномірність завантаження вузлів інфраструктури призводить до їх простою, таким чином виникає потреба у впровадженні технологій, які не призводять до простоїв обладнання, а також гарантують якість обслуговування навантаження протягом дня.

2. Визначено, що для задоволення ряду суперечливих вимог до мережеских функцій необхідно застосовувати гібридні системи, які дозволяють спільно використовувати технології апаратної та віртуальної реалізації мережевої функції.

3. Розроблено метод визначення місця розміщення та обсягів зарезервованих обчислювальних ресурсів віртуальних мережеских функцій у дата центрах оператора мобільного зв'язку, який гарантує якість надання телекомунікаційних сервісів з мінімально необхідними затратами ресурсів, виконуючи визначення їх достатньої конфігурації у гібридному середовищі доступних ресурсів, що дозволяє скоротити витрати на 13% у порівнянні з випадково обраною монохмарою та на 47% у порівнянні з традиційним підходом до розгортання мережі.

4. Розроблено метод визначення розміру інтервалу часу сталої конфігурації обчислювальних ресурсів, який передбачає його змінну величину і врахування як витрат на реконфігурацію, так і коефіцієнту використання ресурсів, забезпечуючи гнучке використання ресурсів у віртуалізованому середовищі, зменшуючи відсоток вільних ресурсів на 42% у порівнянні з виділеним обладнанням і на 9% у порівнянні з існуючими аналогами та зменшуючи робоче навантаження у мережі.

5. Удосконалено розподілений метод локальної реконфігурації обчислювальних ресурсів віртуальної мережі у випадку відмови або перевантаження, який за рахунок децентралізованого керування та врахування міграційних витрат перерозподіляє віртуальні мережеві функції в нормальному та нештатному режимі із забезпеченням економічно обґрунтованого використання ресурсів, зменшуючи витрати в середньому на 21%.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Bradai A. Cellular software defined networking: a framework / A. Bradai, K. Singh, T. Ahmed, T. Rasheed // IEEE Communications Magazine. – 2015. – Vol. 53, No. 6. – P. 36-43.
2. Basta A. Applying NFV and SDN to LTE mobile core gateways, the functions placement problem / A. Basta, W. Kellerer, M. Hoffmann, H. Morper et al. // 4th workshop on All things cellular: operations, applications, & challenges. – Chicago, USA, 2014. – P. 33-38.
3. Yousaf F. Z. SoftEPC – Dynamic instantiation of mobile core network entities for efficient resource utilization / F. Z. Yousaf, J. Lessmann, P. Loureiro, S. Schmid // 2013 IEEE International Conference on Communications (ICC). – Budapest, Hungary, 2013. – P. 3602-3606.
4. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2016–2021 White Paper [Online]. – Available at: <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf>
5. Kopetz H. Internet of Things. In: Real-Time Systems. – Springer, Boston, MA, 2011. – P. 307-323.
6. 5G White paper [Online]. – Available at: https://www.nttdocomo.co.jp/english/binary/pdf/corporate/technology/whitepaper_5g/DOCOMO_5G_White_Paper.pdf
7. The Internet of everything – A \$19 trillion opportunity [Online]. – Available at: <http://www.cisco.com/web/services/portfolio/consulting-services/documents/consulting-services-capturing-ioe-valueaag.pdf>
8. Emmerson B. M2M: The internet of 50 billion devices / B. Emmerson // WinWin Magazine. – 2010. – P. 19-22.
9. Shimojo T. Future mobile core network for efficient service operation / T. Shimojo, Y. Takano, A. Khan, S. Kaptchouang, M. Tamura, S. Iwashina // 1st

IEEE Conference on Network Softwarization (NetSoft). – London, UK, 2015.– P. 1-6.

10. Shimojo T. Cost-efficient method for managing network slices in a multi-service 5G core network / T. Shimojo, M. R. Sama, A. Khan, S. Iwashina // Proceedings of the 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM). – Lisbon, Portugal, 2017. – P. 1121-1126.

11. Sama M. R. Reshaping the mobile core network via function decomposition and network slicing for the 5G Era / M. R. Sama, X. An, Q. Wei, S. Beker // Proceedings of the 2016 IEEE Wireless Communications and Networking Conference Workshops (WCNCW). – Doha, Qatar, 2016. – P. 90-96.

12. Mijumbi R. Network Function Virtualization: State-of-the-art and Research Challenges / R. Mijumbi, J. Serrat, J. Gorricho, N. Bouten et al. // IEEE Communications Surveys & Tutorials. – 2015. – Vol. 18, No. 1. – P. 236-262.

13. Hawilo H. NFV: state of the art, challenges, and implementation in next generation mobile networks (vEPC) / H. Hawilo, A. Shami, M. Mirahmadi, R. Asal // IEEE Network. – 2014. – Vol. 28, No. 6. – P. 18-26.

14. Liquid Core [Online]. – Available at: <http://networks.nokia.com/portfolio/liquidnet/liquidcore>.

15. Soares J. Cloud4nfv: A platform for virtual network functions / J. Soares, M. Dias, J. Carapinha, B. Parreira, S. Sargento // 2014 IEEE 3rd International Conference on Cloud Networking (CloudNet). – Luxembourg, 2014. – P. 288-293.

16. Moens H. VNF-P: A model for efficient placement of virtualized network functions / H. Moens, F. De Turck // 10th International Conference on Network and Service Management. – Rio de Janeiro, 2014. – P. 418-423.

17. Future Communication Architecture For Mobile Cloud Services. Overall Architecture Definition [Online]. – Available at: <http://www.mobile-cloud-networking.eu/site/index.php?process=download&id=124&code=93b79f8f5b99f67a6cdc28369c05b65f624cfee7>

18. ETSI Industry Specification Group (ISG), "Draft Documents." Aug2015 [Online]. – Available at: <https://docbox.etsi.org/isg/nfv/open/Drafts/>
19. Signaling is growing 50% faster than data traffic [Online]. – Available at: <http://docplayer.net/6278117-Signaling-is-growing-50-faster-than-data-traffic.html>
20. Mueller J. Mobile Cloud – Combining EPC, SDN and NFV / J. Mueller, T. Magendaz // Mobile Network (Function) Virtualization and Software Defined Networks of ITG Informationstechnische Gesellschaft im VDE FG 5.2.4 'Mobility in IP-based Networks'. – Munich, Germany, 2013.
21. Mijumbi R. On the energy efficiency prospects of network function virtualization / R. Mijumbi, J. Serrat, J. Gorricho and J. Rubio-Loyola [Online]. – Available at: <https://arxiv.org/pdf/1512.00215.pdf>
22. Network Functions Virtualisation (NFV); Use Cases [Online]. – Available at: http://www.etsi.org/deliver/etsi_gs/nfv/001_099/001/01.01.01_60/gs_nfv001v010101p.pdf.
23. The Rise of Virtual EPC: A Mobile Packet Core Forecast & Analysis [Online]. – Available at: http://www.heavyreading.com/details.asp?sku_id=3187&skuitem_itemid=1558
24. Economic Benefits of Virtualized Evolved Packet Core [Online]. – Available at: <https://www.cisco.com/c/dam/en/us/solutions/collateral/service-provider/ultra-services-platform/idc-us41419916.pdf>
25. The LTE Network Architecture [Online]. – Available at: http://www.cse.unt.edu/~rdantu/FALL_2013_WIRELESS_NETWORKS/LTE_Alc_atel_White_Paper.pdf
26. 3GPP Technical Specification 23.203, Policy and charging control architecture [Online]. – Available at: http://www.etsi.org/deliver/etsi_ts/123200_123299/123203/12.06.00_60/ts_123203v120600p.pdf

27. Sesia S. LTE – The UMTS Long Term Evolution: From Theory to Practice / S. Sesia, I. Toufik, M. Baker. – Wiley, 2009. – 792 p.
28. Herrera J. G. Resource Allocation in NFV: A Comprehensive Survey / J. G. Herrera, J. F. Botero // IEEE Transactions on Network and Service Management. – 2016. – Vol. 13, No. 3. – P. 518-532.
29. Network Functions Virtualisation (NFV); Architectural Framework [Online]. – Available at: http://www.etsi.org/deliver/etsi_gs/nfv/001_099/002/01.01.01_60/gs_nfv002v010101p.pdf.
30. ETSI NFV Management & Orchestration (MANO) простым языком [Електронний ресурс] : [Веб-сайт]. – Електронні дані. – Режим доступа: <https://sdnblog.ru/etsi-nfv-mano-beginners-tutorial/>. – Назва з екрана.
31. Скулиш М. А. Метод складання розкладу залучення ресурсів для високонавантажених інформаційних систем / М. А. Скулиш // Наукові записки Українського науково-дослідного інститут зв'язку. – 2014. – № 6. – С. 65-70.
32. Sitaram D. Moving To The Cloud: Developing Apps in the New World of Cloud Computing / D. Sitaram, G. Manjunath. – Massachusetts: Syngress, 2011. – 468 p.
33. Vu H.T. A Traffic and Power-aware Algorithm for Virtual Machine Placement in Cloud Data Center / H.T. Vu, S. Hwang // International Journal of Grid and Distributed Computing. – 2014. – Vol. 7, No. 1. – pp. 21-32.
34. M. Peter and G. Timothy. The NIST Definition of Cloud Computing, Recommendations of the National Institute of Standards and Technology [Online]. – Available at: <http://www.nist.gov/itl/cloud/>.
35. Scharf M. Monitoring and abstraction for networked clouds / M. Scharf, T. Voith, W. Roome, B. Gaglianella, M. Steiner, V. Hilt, and V. Gurbani // 2012 16th International Conference on Intelligence in Next Generation Networks (ICIN). – Berlin, Germany, 2012. – P. 80-85.

36. Mandal A. Provisioning and Evaluating Multi-domain Networked Clouds for Hadoop-based Applications / A. Mandal, Y. Xin, I. Baldine, P. Ruth, C. Heerman, J. Chase, V. Orlikowski, and A. Yumerefendi // 2011 IEEE Third International Conference on Cloud Computing Technology and Science (CloudCom). – Athens, Greece, 2011. – P. 690-697.
37. Lopez D. Network functions virtualization: Beyond carrier-grade clouds / D. Lopez // Optical Fiber Communications Conference and Exhibition (OFC). – San Francisco, CA, USA, 2014. – P. 1-18.
38. H. Basilier M. Darula and J. Wilke. Virtualizing network services the telecom cloud. Technical White Paper [Online]. – Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.488.5158&rep=rep1&type=pdf>
39. Distributed Management Task Force, “Cloud Management Working Group (CMWG),” [Online]. – Available at: <http://www.dmtf.org/standards/cmwg>
40. McKeown N. OpenFlow: Enabling Innovation in Campus Networks / N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner // SIGCOMM Comput. Commun. Rev. – 2008. – Vol. 38, No. 2. – pp. 69-74.
41. Mauro D.R. Essential SNMP / D. R. Mauro and K. J. Schmidt. – O'Reilly Media, 2005. – 464 p.
42. Wedge G. Carrier-Grade: Five Nines, the Myth and the Reality / G. Wedge and L. Barbara // Pipeline Publications. – 2006. – Vol. 3, No. 1. – P. 1-14.
43. Vishwanath K. V. Characterizing Cloud Computing Hardware Reliability / K. V. Vishwanath and N. Nagappan // ACM Symposium on Cloud Computing. – New York, NY, USA, 2010. – P. 193-204.
44. Network Equipment-Building System (NEBS) [Online]. – Available at: http://en.wikipedia.org/wiki/Network_Equipment-Building_System
45. Adamuz-Hinojosa O. Automated Network Service Scaling in NFV: Concepts, Mechanisms and Scaling Workflow / O. Adamuz-Hinojosa, J. Ordonez-

Lucena, P. Ameigeiras, J. J. Ramos-Munoz, D. Lopez, and J. Folgueira // IEEE Communications Magazine. – 2018. – Vol. 56, No. 7. – P. 162-169.

46. Network Functions Virtualisation (NFV); Management and Orchestration [Online]. – Available at: https://www.etsi.org/deliver/etsi_gs/NFV-MAN/001_099/001/01.01.01_60/gs_NFV-MAN001v010101p.pdf.

47. Network Functions Virtualisation (NFV); Virtual Network Functions Architecture [Online]. – Available at: https://www.etsi.org/deliver/etsi_gs/NFV-SWA/001_099/001/01.01.01_60/gs_NFV-SWA001v010101p.pdf.

48. Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; Network Service Templates Specification [Online]. – Available at: https://www.etsi.org/deliver/etsi_gs/NFV-IFA/001_099/014/02.04.01_60/gs_nfv-ifa014v020401p.pdf.

49. Cheng X. Virtual network embedding through topology-aware node ranking / X. Cheng, S. Su, Z. Zhang, H. Wang, F. Yang, Y. Luo, J. Wang // ACM SIGCOMM Computer Communication Review. – 2011. – Vol. 41, No. 2. – P. 38-47.

50. Zhang S. Virtual network embedding with substrate support for parallelization / S. Zhang, J. Wu, and S. Lu // IEEE GLOBECOM. – Anaheim, CA, USA, 2012. – P. 2615-2620.

51. Gandhi A. Minimizing Data Center SLA Violations and Power Consumption via Hybrid Resource Provisioning / A. Gandhi, Y. Chen, D. Gmach, M. Arlitt, and M. Marwah // Green Computing : International Conference and Workshops, Orlando, FL, 25–28 July 2011 : proceedings. – IEEE, 2011. – P. 1–8.

52. Xu Z. Efficient virtual network embedding via exploring periodic resource demands / Z. Xu, W. Liang, Q. Xia // 2014 IEEE 39th Conference on Local Computer Networks (LCN). – Edmonton, AB, Canada, 2014. – P. 90-98.

53. Beloglazov A. Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in Cloud data centers / A. Beloglazov, R. Buyya // Concurrency and Computation: Practice & Experience. – 2012. – Vol. 24, No. 13. – P. 1397-1420.

54. Gandhi A. Server farms with setup costs / A. Gandhi, M. Harchol-Balter, I. Adan // Performance Evaluation. – 2010. – Vol. 67. – P. 1123-1138.
55. Coskun A. Evaluating the impact of job scheduling and power management on processor lifetime for chip multiprocessors / A. Coskun, R. Strong, D. Tullsen, S. Rosing // ACM SIGMETRICS. – WA, USA, 2009. – P. 169-180.
56. Khuller S. Energy efficient scheduling via partial shutdown / S. Khuller , J. Li , B. Saha // Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms. – Texas, USA, 2010. – pp. 1360-1372.
57. Rajan A.S. Understanding the bottlenecks in virtualizing cellular core network functions / A.S. Rajan, S. Gobriel, C. Maciocco, K.B. Ramia et al. // 2015 IEEE International Workshop on Local and Metropolitan Area Networks (LANMAN). – Beijing, 2015. – P. 1-6.
58. Ferrer Riera J. Virtual network function scheduling: Concept and challenges / J. Ferrer Riera, E. Escalona, J. Batlle, E. Grasa, J. A. Garcia-Espin // 2014 International Conference on Smart Communications in Network Technologies (SaCoNeT). – Vilanova i la Geltru, 2014. – pp. 1-5.
59. Jennings B. Resource management in clouds: Survey and research challenges / B Jennings, R Stadler // Journal of Network and Systems Management. – 2014. – P. 1-53.
60. Fischer A. Virtual Network Embedding: A Survey / A. Fischer, J. Botero, M. Till Beck, H. de Meer and X. Hesselbach // IEEE Communications Surveys & Tutorials. – 2013. – Vol. 15, No. 4. – P. 1888-1906.
61. Baumgartner A. Mobile core network virtualization: A model for combined virtual core network function placement and topology optimization / A. Baumgartner, V.S. Reddy, T. Bauschert // 2015 1st IEEE Conference on Network Software-ization (NetSoft). – London, 2015. – P. 1-9.
62. Mehraghdam S. Specifying and placing chains of virtual network functions / S. Mehraghdam, M. Keller, H. Karl // 2014 IEEE 3rd International Conference on Cloud Network-ing (CloudNet). – Luxembourg, 2014. – P. 7-13.

63. Lischka J. A virtual network mapping algorithm based on subgraph isomorphism detection / J. Lischka, H. Karl // Proceedings of the 1st ACM workshop on Virtualized infra-structure systems and architectures. – 2009. – P. 81-88.
64. Heller B. The controller placement problem / B. Heller, R. Sherwood, N. McKeown // ACM HotSDN. – Helsinki, Fin-land, 2012. – P. 1-6.
65. Hock D. Pareto-optimal resilient controller placement in SDN-based core networks / D. Hock, M. Hartmann, S. Gebert, M. Jarschel et al. // 25th International Teletraffic Congress (ITC). – Shanghai, 2013. – P. 1-9.
66. Lange S. Heuristic Approaches to the Controller Placement Problem in Large Scale SDN Networks / S. Lange, S. Gebert, T. Zinner, P. Tran-Gia et al. // IEEE Transactions on Network and Service Management. – 2015. – Vol. 12, No. 1. – P. 4-17.
67. Chandra A. Dynamic resource allocation for shared data centers using online measurements / A. Chandra, W. Gong, P. Shenoy // 11th international conference on Quality of service. – Berkeley, CA, USA, 2003. – P. 381-398.
68. Infuhr J. Introducing the virtual network mapping problem with delay, routing and location constraints / J. Infuhr, and G. R. Raidl // 5th international conference on Network optimization. – Springer-Verlag, 2011. – P. 105-117.
69. Jarray A. VCG auction-based approach for efficient Virtual Network embedding / A. Jarray and A. Karmouch // IFIP/IEEE International Symposium on Integrated Network Management. – Ghent, Belgium, 2013. – P. 609-615.
70. Cai Z. Virtual network embedding for evolving networks / Z. Cai, F. Liu, N. Xiao, Q. Liu, and Z. Wang // IEEE Global Telecommunications Conference. – Miami, FL, USA, 2010. – P. 1-5.
71. Sun G. A cost efficient framework and algorithm for embedding dynamic virtual network requests / G. Sun, H. Yu, V. Anand, L. Li // Future Generation Computer Systems. – 2013. – Vol. 29, No. 5. – P. 1265-1277.
72. Mijumbi R. Design and Evaluation of Learning Algorithms for Dynamic Resource Management in Virtual Networks / R. Mijumbi, J.-L. Gorricho,

J. Serrat, M. Claeysy, F. D. Turkey, S. Latr // Network Operations and Management Symposium (NOMS), Krakow, 5–9 May 2014. – IEEE, 2014. – P. 1–9.

73. Patikirikoral T. A multi-model framework to implement self-managing control systems for QoS management / T. Patikirikoral, A. Colman, J. Han, and L. Wang // 6th International Symposium on Software Engineering for Adaptive and Self-Managing Systems. – Waikiki, Honolulu, HI, USA, 2011. – P. 218-227.

74. Lai W. Game Theoretic Distributed Dynamic Resource Allocation with Interference Avoidance in Cognitive Femtocell Networks / W. Lai, M. Chiang, S. Lee, T. Lee // IEEE Wireless Communications and Networking Conference. – Shanghai, China, 2013. – P. 3364-3369.

75. Jokhio F. Prediction-Based Dynamic Resource Allocation for Video Transcoding in Cloud Computing / F. Jokhio, A. Ashraf, S. Lafond, I. Porres, J. Lilius // 21st Euromicro International Conference on Parallel, Distributed, and Network-Based Processing. – Belfast, UK, 2012. – P. 254-261.

76. Castellanos M. iBOM: A Platform for Intelligent Business Operation Management / M. Castellanos, F. Casati, M. Shan, U. Dayal // Int. Conf. on Data Engineering (ICDE). – Tokyo, Japan, 2005. – P. 1084-1095.

77. Xu J. Autonomic resource management in virtualized data centers using fuzzy logic based approaches / J. Xu, M. Zhao, J. Fortes, R. Carpenter, M. Yousif // Cluster Computing Journal. – 2008. – Vol. 11, No. 3. – P.213-227.

78. Urgaonkar B. Agile dynamic provisioning of multi-tier Internet applications / B. Urgaonkar, P. Shenoy, A. Chandra, P. Goyal, T. Wood // ACM Transactions on Autonomous and Adaptive Systems (TAAS). – 2008. – Vol. 3, No. 1. – P.1-39.

79. Коваль Б. В. Дослідження площини управління програмно-керованих мереж на основі розподіленої системи функцій віртуалізації / Б. В. Коваль, М. О. Селюченко, Г. В. Мельник, А. В. Ковальчук // Вісник Національного університету «Львівська політехніка». Радіoeлектроніка та телекомунікації. – 2014. – № 796. – С. 164-175.

80. Xiong G. A virtual service placement approach based on improved quantum genetic algorithm / G. Xiong, Y.-X. Hu, L. Tian, J.-L. Lan, J.-F. Li, And Q. Zhou // *Frontiers of Information Technology & Electronic Engineering*. – 2016. – Vol. 17, No. 7. – P. 661–671.
81. Gember-Jacobson A. OpenNF: Enabling Innovation in Network Function Control / A. Gember-Jacobson, R. Viswanathan, C. Prakash, R. Grandl, J. Khalid, S. Das, and A. Akella // *Proceedings of the 2014 ACM Conference on SIGCOMM*, ser. SIGCOMM '14. – New York, NY, USA, 2014. – P. 163-174.
82. Rajagopalan S. Split/Merge: System Support for Elastic Execution in Virtual Middleboxes / S. Rajagopalan, D. Williams, H. Jamjoom, and A. Wareld // *10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)*. – Lombard, IL, USA, 2013. – P. 227-240.
83. Xia J. Optimized Virtual Network Functions Migration for NFV/ J. Xia, Z. Cai, M. Xu // *IEEE 22nd International Conference on Parallel and Distributed Systems*. – Wuhan, China, 2016. – P. 340-346.
84. МСЭ-Т E.800 [Электронный ресурс]. – Режим доступа: https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-E.800-200809-I!!PDF-R&type=items.
85. Scholler M. Resilient deployment of virtual network functions / M. Scholler, M. Stiernerling, A. Ripke, R. Bless // *Proc. 5th Int. Congr. Ultra Mod. Telecommun. Control Syst. Workshops (ICUMT)*. – St. Petersburg, Russia, 2013. – P. 208-214.
86. Taleb T. On Service Resilience in Cloud-Native 5G Mobile Systems / T. Taleb, A. Ksentini, B. Sericola // *IEEE Journal on Selected Areas in Communications*. – 2016. – Vol. 34, No. 3. – P. 483-496.
87. Chowdhury N.M.M.K. Virtual Network Embedding with Coordinated Node and Link Mapping / N.M.M.K. Chowdhury, M. Rahman and R. Boutaba // *INFCOM*. – Rio de Janeiro, Brazil, 2009. – P. 1-9.
88. Fajjari I. VNR Algorithm: A Greedy Approach For Virtual Networks Reconfigurations / I. Fajjari, N. Aitsaadi, G. Pujolle and H. Zimmermann. // *IEEE*

Global Communications Conference, Exhibition and Industry Forum. – Houston, United States, 2011. – P. 1-7.

89. Rahman M.R. Survivable Virtual Network Embedding / M. R. Rahman, I. Aib, and R. Boutaba // NETWORKING 2010. – 2010. – P. 40-52.

90. Abid H. A novel scheme for node failure recovery in virtualized networks / H. Abid; N. Samaan // 2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013). – Ghent, Belgium, 2013. – P. 1154-1160.

91. Скулиш М. А. Організація управління в багатоступінних системах масового обслуговування / М. А. Скулиш, С. В. Суліма // Вісник Національного університету «Львівська політехніка». Радіoeлектроніка та телекомунікації. — 2015. — № 818. — С. 80–85.

92. Skulysh M. Management of Multiple Stage Queuing Systems / M. Skulysh, S. Sulima // CADSM 2015 : 13-th International conference, 24–27 February 2015 : IEEE Digital Library. — Lviv–Polyana, 2015. — P. 431–433.

93. Скулиш М. А. Організація управління в багатоступінних системах масового обслуговування / М. А. Скулиш, С. В. Суліма // Проблеми телекомунікацій ПТ-2015 : 9-а міжнародна науково-технічна конференція, 21–24 квітня 2015 : матеріали конференції. — Київ, 2015. — С. 319–321.

94. Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); General Packet Radio Service (GPRS); Service description; Stage 2 [Online]. – Available at: http://www.etsi.org/deliver/etsi_ts/123000_123099/123060/10.03.00_60/ts_123060_v100300p.pdf

95. LTE; General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access [Online]. – Available at: http://www.etsi.org/deliver/etsi_ts/123400_123499/123401/12.06.00_60/ts_123401_v120600p.pdf

96. Universal Mobile Telecommunications System (UMTS); LTE; Architecture enhancements for non-3GPP accesses [Online]. – Available at: http://www.etsi.org/deliver/etsi_ts/123400_123499/123402/10.04.00_60/ts_123402v100400p.pdf
97. Olsson M. EPC and 4G Packet Networks / M. Olsson, C. Mulligan. – Academic Press, 2012. – 624 p.
98. Prados-Garzon J. Modeling and Dimensioning of a Virtualized MME for 5G Mobile Networks / J. Prados-Garzon, J. J. Ramos-Munoz, P. Ameigeiras, P. Andres-Maldonado, J. M. Lopez-Soler // IEEE Transactions on Vehicular Technology. – 2017. – Vol. 66, No. 5. – P. 4383-4395.
99. Mijumbi R. Server placement and assignment in virtualized radio access networks / R. Mijumbi, J. Serrat, J.-L. Gorricho, J. Rubio-Loyola, S. Davy // 2015 11th International Conference on Network and Service Management (CNSM). – Barcelona, 2015. – P. 398-401.
100. Cau E. Efficient Exploitation of Mobile Edge Computing for Virtualized 5G in EPC Architectures / E. Cau, M. Corici, P. Bellavista, L. Foschini, G. Carella, A. Edmonds, T. M. Bohnert // 2016 4th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering. – Oxford, UK, 2016. – P. 100-109.
101. Network time protocol version 4: Protocol and algorithms specification [Online]. – Available at: <https://tools.ietf.org/html/rfc5905>
102. Jeon S. Virtualised EPC for on-demand mobile traffic offloading in 5G environments / S. Jeon, D. Corujo, R. L. Aguiar // 2015 IEEE Conference on Standards for Communications and Networking (CSCN). – Tokyo, Japan, 2016. – P. 275-281.
103. Фелижанко А. Виртуализация в сетях мобильной связи [Электронный ресурс]. – Режим доступа: <https://www.slideshare.net/CiscoRu/celc-virtualization-inmobilenetworks17042014afelizha>

104. Математичні методи аналізу та керування телекомунікаційними мережами. Розділ 5 // Л.С. Глоба, О.М. Дяденко, А.Ю. Пилипенко, М.А. Скулиш. –К.: Інститут обдарованої дитини НАПН України. — ISBN 978-966-2633-84-9. — 2018 . — С. 206-225.

105. Skulysh M. Model for Efficient Allocation of Network Functions in Hybrid Environment / M. Skulysh, L. Globa, S. Sulima // Information and Telecommunication Sciences. — 2016. — № 1. — С. 39–45.

106. Скулиш М. А. Метод управління ресурсами в датацентрах оператора мережі мобільного зв'язку / М. А. Скулиш, С. В. Суліма // Радіоелектроніка та інформатика. — 2015. — № 3. — С. 8–13.

107. Скулиш М. А. Управління ресурсами в віртуалізованих мережах оператора мобільного зв'язку / М. А. Скулиш, С. В. Суліма // Інфокомунікації – сучасність та майбутнє : 5-а міжнародна науково-практична конференція, 29-30 жовтня 2015 : матеріали конференції. — Одеса, 2015. — С.84–88.

108. Globa L. Method for resource allocation of virtualized network functions in hybrid environment / L. Globa, M. Skulysh, S. Sulima // 2016 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), 6–9 June 2016 : IEEE Digital Library. — Varna, Bulgaria, 2016. — P. 1–5.

109. Skulysh M. Service deployment aspects in the systems with network function virtualization / M. Skulysh, S. Sulima // Radio Electronics & Info Communications : First International Conference, 11–16 September 2016 : IEEE Digital Library. — Kyiv, 2016. — P. 1–7.

110. Sunduchkov K. Sampling for direct search method of all system implementations developed according to the requirements of the technical design specification / K. Sunduchkov, M. Skulysh, S. Sulima, B. Savchuk // 2017 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), 10–13 October 2017 : IEEE Digital Library. — Kharkov, 2017. — P. 83–86.

111. Суліма С. В. Метод відновлення вузла у мережах NFV / С. В. Суліма // Проблеми телекомунікацій ПТ-2017 : 11-а міжнародна науково-технічна конференція, 18–21квітня 2017 : матеріали конференції. — Київ, 2017 . — С. 270–272.
112. Bianco A. Cost and performance trade-offs in reconfiguration strategies for WDM networks / A. Bianco, J. Finochietto, C. Piglione// 4th International Telecommunication Networking Workshop on QoS in Multiservice IP Networks. – Venice, Italy, 2008. – P. 95-100.
113. Chen Y. Managing Server Energy and Operational Costs in Hosting Centers / Y. Chen, A. Das, W. Qin, A. Sivasubramaniam, Q. Wang, N. Gautam // SIGMETRICS '05 Proceedings of the 2005 ACM SIGMETRICS international conference on Measurement and modeling of computer systems. – Banff, Alberta, Canada, 2005. – P. 303-314.
114. Экспоненциальное сглаживание [Электронный ресурс] : [Вэб-сайт]. — Электронные данные. — Режим доступа: <http://www.machinelearning.ru>. — Название с экрана.
115. Skulysh M. Hybrid resource management system for telecommunication network / M. Skulysh, S. Sulima // Advanced Information Systems. — 2018. — Vol. 2, № 1. — P. 47–51.
116. Суліма С. В. Система управління ресурсами в центрах обробки даних оператора мережі мобільного зв'язку / С. В. Суліма, М. А. Скулиш // Вісник НТУУ «КПІ». Серія Радіотехніка, Радіоапаратобудування. — 2017 . — № 68. — С. 27-32. (Наукове фахове видання України з технічних наук. Входить до міжнародних наукометричних баз: DOAJ, Web of Science, Index Copernicus, eLibrary.ru / РИНЦ та ін.)
117. Суліма С. В. Гібридна система управління ресурсами для віртуалізованих мережових функцій / С. В. Суліма, М. А. Скулиш // Радіoeлектроніка, інформатика, управління.— 2017 . — № 1(40). — С. 16–23. (Наукове фахове видання України з технічних наук. Входить до міжнародних

наукометричних баз: Web of Science, DOI, DOAJ, British Library, eLibrary.ru / РИНЦ, Index Copernicus та ін.)

118. Суліма С. В. Алгоритм відображення та планування віртуалізованих функцій в мережі мобільного зв'язку / С. В. Суліма, М. А. Скулиш // Проблеми телекомунікацій ПТ-2016 : 10-а міжнародна науково-технічна конференція, 19–22 квітня 2016 : матеріали конференції. — Київ, 2016. — С. 372–374.

119. Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; Or-Vnfm Reference Point - Interface and Information Model Specification [Online]. – Available at: https://www.etsi.org/deliver/etsi_gs/NFV-IFA/001_099/005/02.04.01_60/gs_nfv-ifa005v020401p.pdf.

120. Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; Or-Vi reference point - Interface and Information Model Specification [Online]. – https://www.etsi.org/deliver/etsi_gs/NFV-IFA/001_099/007/02.04.01_60/gs_nfv-ifa007v020401p.pdf.

121. Суліма С. В. Метод відновлення мережі у віртуалізованому середовищі / С. В. Суліма // Радіoeлектроніка та інформатика. — 2017. — № 4. — С. 4–8.

122. Суліма С. В. Метод відновлення вузла у мережах NFV / С. В. Суліма // Проблеми телекомунікацій ПТ-2017 : 11-а міжнародна науково-технічна конференція, 18–21 квітня 2017 : матеріали конференції. — Київ, 2017. — С. 270–272.

ДОДАТОК А. СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ ТА ВІДОМОСТІ ПРО АПРОБАЦІЮ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЇ

Монографія

1. Математичні методи аналізу та керування телекомунікаційними мережами. Розділ 5 // Л.С. Глоба, О.М. Дяденко, А.Ю. Пилипенко, М.А. Скулиш. –К.: Інститут обдарованої дитини НАПН України. — ISBN 978-966-2633-84-9. — 2018 . — С. 206-225.

Особистий внесок: система керування відображенням та плануванням ресурсів телекомунікаційної мережі з використанням технології віртуалізації.

Статті у наукових фахових виданнях

2. Скулиш М. А. Організація управління в багатоетапних системах масового обслуговування / М. А. Скулиш, С. В. Суліма // Вісник Національного університету «Львівська політехніка». Радіoeлектроніка та телекомунікації. — 2015. — № 818. — С. 80–85.

Особистий внесок: аналітична модель для визначення кількості ресурсів, яку необхідно виділити для кожного етапу багатоетапної системи.

3. Skulysh M. Model for Efficient Allocation of Network Functions in Hybrid Environment / M. Skulysh, L. Globa, S. Sulima // Information and Telecommunication Sciences. — 2016. — № 1. — С. 39–45.

Особистий внесок: модель виділення ресурсів системи даних центрів для визначення оптимальної кількості ресурсів виділених мережевому функціональному вузлу.

4. Суліма С. В. Метод відновлення мережі у віртуалізованому середовищі / С. В. Суліма // Радіoeлектроніка та інформатика. — 2017. — № 4. — С. 4–8.

5. Skulysh M. Hybrid resource management system for telecommunication network / M. Skulysh, S. Sulima // Advanced Information Systems. — 2018. — Vol. 2, № 1. — P. 47–51.

Особистий внесок: спосіб керування ресурсами телекомунікаційної мережі шляхом вибору оптимального інтервалу для перерозподілу системних ресурсів.

Статті у фахових виданнях України, які включені до міжнародних наукометричних баз

6. Суліма С. В. Система управління ресурсами в центрах обробки даних оператора мережі мобільного зв'язку / С. В. Суліма, М. А. Скулиш // Вісник НТУУ «КПІ». Серія Радіотехніка, Радіоапаратобудування. — 2017. — № 68. — С. 27-32. (Наукове фахове видання України з технічних наук. Входить до міжнародних наукометричних баз: DOAJ, Web of Science, Index Copernicus, eLibrary.ru / РИНЦ та ін.)

Особистий внесок: метод динамічної конфігурації ресурсів.

7. Суліма С. В. Гібридна система управління ресурсами для віртуалізованих мережевих функцій / С. В. Суліма, М. А. Скулиш // Радіоелектроніка, інформатика, управління.— 2017. — № 1(40). — С. 16–23. (Наукове фахове видання України з технічних наук. Входить до міжнародних наукометричних баз: Web of Science, DOI, DOAJ, British Library, eLibrary.ru / РИНЦ, Index Copernicus та ін.)

Особистий внесок: метод побудови системи управління ресурсами для віртуальних мережевих функцій в центрах обробки даних оператора мережі мобільного зв'язку.

Статті у інших виданнях

8. Скулиш М. А. Метод управління ресурсами в датацентрах оператора мережі мобільного зв'язку / М. А. Скулиш, С. В. Суліма // Радіоелектроніка та інформатика. — 2015. — № 3. — С. 8–13.

Особистий внесок: модель для визначення оптимальної кількості датацентрів, топології їх розміщення та кількості ресурсів, які виділяються

функціональним блокам системи, що розгортаються в дата центрах та передбачають використання декількох типів ресурсів.

Тези наукових доповідей

9. Skulysh M. Management of Multiple Stage Queuing Systems / M. Skulysh, S. Sulima // CADSM 2015 : 13-th International conference, 24–27 February 2015 : IEEE Digital Library. — Lviv–Polyana, 2015. — P. 431–433.

Особистий внесок: модель динамічного виділення ресурсів у багатоступінних системах.

10. Скулиш М. А. Організація управління в багатоступінних системах масового обслуговування / М. А. Скулиш, С. В. Суліма // Проблеми телекомунікацій ПТ-2015 : 9-а міжнародна науково-технічна конференція, 21–24 квітня 2015 : матеріали конференції. — Київ, 2015. — С. 319–321.

Особистий внесок: метод динамічного керування ресурсами багатоступінних систем.

11. Скулиш М. А. Управління ресурсами в віртуалізованих мережах оператора мобільного зв'язку / М. А. Скулиш, С. В. Суліма // Інфокомунікації – сучасність та майбутнє : 5-а міжнародна науково-практична конференція, 29-30 жовтня 2015 : матеріали конференції. — Одеса, 2015. — С. 84–88.

Особистий внесок: метод початкової організації гетерогенних ресурсів мережі.

12. Globa L. Method for resource allocation of virtualized network functions in hybrid environment / L. Globa, M. Skulysh, S. Sulima // 2016 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), 6–9 June 2016 : IEEE Digital Library. — Varna, Bulgaria, 2016. — P. 1–5.

Особистий внесок: метод адаптивної конфігурації гетерогенних ресурсів мережі.

13. Суліма С. В. Алгоритм відображення та планування віртуалізованих функцій в мережі мобільного зв'язку / С. В. Суліма, М. А. Скулиш // Проблеми телекомунікацій ПТ-2016 : 10-а міжнародна науково-технічна

конференція, 19–22 квітня 2016 : матеріали конференції. — Київ, 2016. — С. 372–374.

Особистий внесок: метод відображення та планування віртуалізованих мережесевих функцій.

14. Skulysh M. Service deployment aspects in the systems with network function virtualization / M. Skulysh, S. Sulima // Radio Electronics & Info Communications : First International Conference, 11–16 September 2016 : IEEE Digital Library. — Kyiv, 2016. — P. 1–7.

Особистий внесок: підхід до моделювання та управління розподілом ресурсів віртуалізованих мережесевих функцій у мережі телекомунікаційного оператора.

15. Sunduchkov K. Sampling for direct search method of all system implementations developed according to the requirements of the technical design specification / K. Sunduchkov, M. Skulysh, S. Sulima, B. Savchuk // 2017 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), 10–13 October 2017 : IEEE Digital Library. — Kharkov, 2017. — P. 83–86.

Особистий внесок: формалізація модифікованого методу оптимізації побудови системи зв'язку.

16. Суліма С. В. Метод відновлення вузла у мережах NFV / С. В. Суліма // Проблеми телекомунікацій ПТ-2017 : 11-а міжнародна науково-технічна конференція, 18–21 квітня 2017 : матеріали конференції. — Київ, 2017. — С. 270–272.

17. Sulima S.V. Model for allocation of network slices in heterogeneous environment / S. V. Sulima // Modern challenges in telecommunications : Twelfth International Scientific Conference, 16–20 April 2018 : conference proceedings. — Kyiv, 2018. — С. 42–44.

ДОДАТОК Б. АКТИ ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОЇ РОБОТИ

"ЗАТВЕРДЖУЮ"



Начальник департаменту стратегії та розвитку нових напрямків бізнесу

ТОВ «лайфселл»

Лебедєв А.В.

"17" вересня 2018 р.

АКТ

впровадження результатів дисертаційної роботи
Суліми Світлани Валеріївни

Ми, що нижче підписалися, представники Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»:

- завідувач кафедри ІТМ д.т.н., проф. Глоба Л.С.,
- докторант кафедри ІТМ к.т.н., с.н.с. Скулиш М.А.,
- асистент кафедри ІТМ Суліма С.В.

та представники ТОВ «лайфселл»:

- начальник департаменту стратегії та розвитку нових напрямків бізнесу Лебедєв А.В.,
- менеджер проектів з розвитку нових напрямків бізнесу Комардіна О.О.

склали цей акт, який засвідчує, що результати дисертаційної роботи Суліми Світлани Валеріївни, а саме метод визначення розміру інтервалу часу сталої конфігурації обчислювальних ресурсів з його змінною величиною апробовано в лабораторіях ТОВ «лайфселл» під час моделювання роботи ядра мобільної мережі із частковим застосуванням принципу віртуалізації мережевих функцій, що забезпечило гнучке використання ресурсів у віртуалізованому середовищі, доля невикористаних обчислювальних ресурсів зменшився на 32% у порівнянні з обслуговуванням відповідного навантаження на обладнанні фіксованої конфігурації, при цьому показники QoS були збережені на заданому рівні.

ТОВ «лайфселл»
Менеджер проектів з розвитку нових напрямків бізнесу

Комардіна О.О.

(підпис)

17.09.2018

(дата)

КПІ імені Ігоря Сікорського
Зав. кафедри ІТМ, д.т.н., проф.

Глоба Л.С.

Доцент кафедри ІТМ, к.т.н., с.н.с.

Скулиш М.А.

Асистент кафедри ІТМ

Суліма С.В.

13.09.2018

(дата)



УКРАЇНА
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНА УСТАНОВА
НАЦІОНАЛЬНИЙ АНТАРКТИЧНИЙ НАУКОВИЙ ЦЕНТР

01601, м. Київ, бульв. Тараса Шевченка, 16, тел/факс: (380 44) 246-38-80, E-mail: uac@uac.gov.ua www.uac.gov.ua

Від 10.12.18р. № 3
На № _____

ЗАТВЕРДЖУЮ

Т.в.о. директора Державної установи

Національний антарктичний науковий центр

І.В.МОРОЗ

12 2018 р.



впровадження результатів науково-дослідної роботи на тему «Систематизація даних антарктичних досліджень з використанням інтелектуальних засобів представлення знань»

Ми, що нижче підписалися, представники Державної установи Національний антарктичний науковий центр к.б.н. Маланчук В.М., Леонов М.А. та Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» д.т.н., професор Глоба Л.С., к.т.н., н.с. Новогрудська Р.Л., к.т.н., н.с. Скулиш М.А., м.н.с. Суліма С.В. склали цей акт, який засвідчує, що було відпрацьовано програмне забезпечення щодо структуризації та систематизації даних та знань антарктичних досліджень, протоколи підтримки єдиного гетерогенного середовища, яке складається з віртуальних та фізичних ресурсів підтримки функціонування Національного центру Антарктичних даних та вдосконалено програмне забезпечення порталу «Національний центр Антарктичних даних».

Використання моделей та методів систематизації даних антарктичних досліджень дозволило організувати представлення антарктичних даних і знань, а також їх взаємодію в рамках єдиного інформаційно-комунікаційного середовища з передачею результатів антарктичних досліджень на портал NASA. Застосування методу пошуку антарктичних даних, який базується на запропонованій онтологічній моделі та враховує семантичну зв'язність елементів єдиного інформаційно-комунікаційного середовища результатів антарктичних досліджень, дозволив задавати пошуковий запит та інструменти для фільтрації знань у відповідності до метаданих, які характеризують потрібні аспекти наукового дослідження.

Розроблено механізм керування обслуговуванням гібридних сервісів у гетерогенному середовищі для обробки антарктичних даних, а також відповідне програмне забезпечення, яке дозволяє вирішити такі задачі:

- забезпечити постійний контроль за показниками якості обслуговування запитів (затримка у обслуговуванні і ймовірність успішного обслуговування запитів);
- організувати динамічний розподіл обчислювальних ресурсів телекомунікаційної системи, що дозволяє утримувати показники коефіцієнту використання ресурсів у заданих межах.

Під час впровадження роботи використано наукові результати, отримані в дисертаційних роботах к.т.н., н.с. Скулиш М.А., м.н.с. Суліма С.В.

Було відпрацьовано програмно-технічні засоби структуризації та систематизації даних та знань на платформі надання послуг інформування наукової спільноти щодо досліджень, які проводяться в Антарктиці, що дозволило підвищити ефективність роботи із накопиченням, збереженням та обробкою антарктичних даних станції «Академік Вернадський».

Від ДУ НАНЦ

Маланчук В.М.

Леонов М.А.

Від Київ ім. Ігоря Сікорського

Глоба Л.С.

Новогрудська Р.Л.

Скулиш М.А.

Суліма С.В.



ТОВАРИСТВО З ОБМЕЖЕНОЮ ВІДПОВІДАЛЬНІСТЮ «ХУАВЕЙ УКРАЇНА»
LIMITED LIABILITY COMPANY "HUAWEI UKRAINE"

03124Київ, вул. Радищева, буд. 10/14, (ідентифікаційний код: 33501267), тел. 498-77-18, факс 498-77-19
10/14, Radischeva street, Kiev, 03124Ukraine, (Company Code: 33501267), tel. 498-77-18, fax 498-77-19

м. Київ/ Kyiv

Вихідний № 28 від 15.05.2019 р.

"ЗАТВЕРДЖУЮ"
Керівник напрямку Safe & Smart City,
ООО «Хуавей - Україна»

О.М. Дяденко
" 15 " 05 18 р.

АКТ

впровадження результатів дисертаційної роботи
Суліми Світлани Валеріївни

Ми, що нижче підписалися, представники Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» д.т.н., професор Глоба Л.С., к.т.н., доцент Скулиш М.А., Суліма С.В. та керівник напрямку Safe & Smart City, ООО «Хуавей Україна» О.М. Дяденко, менеджер з продукту Safe & Smart City, ООО «Хуавей Україна», склали цей акт про те, що співробітниками КПІ ім. Ігоря Сікорського були розроблені засоби керування перерозподілом обчислювальних ресурсів та телекомунікаційними потоками для вирішення задачі планування ресурсів мобільної мережі з використанням технології віртуалізації.

Розроблений механізм прийняття рішень щодо виділення ресурсів мережі та програмне забезпечення дозволяє вирішити наступні задачі: забезпечити постійний контроль за перевищенням допустимого часу затримки обслуговування службових телекомунікаційних потоків; організувати гнучкий розподіл ресурсів гетерогенної мобільної мережі, що дозволяє підтримувати належний рівень якості обслуговування телекомунікаційних потоків зі змінним трафіком; застосовувати систему керування перерозподілом обчислювальних телекомунікаційних ресурсів, що дозволяє скоротити обсяг незадіяних потужностей обладнання телекомунікаційної мережі.

Проведення апробації запропонованих технічних рішень дозволило покращити такі технічні показники, а саме:

- тестування методу визначення місця розміщення та обсягів зарезервованих обчислювальних ресурсів мережевих функцій гарантує якість надання сервісів з мінімально необхідними витратами ресурсів, визначаючи їх достатню конфігурацію, дозволяє скоротити витрати до 15% у порівнянні з випадково обраною хмарою та на 45-47% у порівнянні з традиційним підходом до розгортання мережі;

- експериментальна перевірка методу локальної реконфігурації обчислювальних ресурсів віртуальної мережі, апробованого у випадку відмови або перевантаження при децентралізованому керуванні та врахуванні міграційних витрат дозволила перерозподілити віртуальні мережеві функції в нормальному та аварійному режимі із забезпеченням економічно обґрунтованого використання обчислювальних ресурсів, зменшити їх витрати в середньому на 17%.

ООО «Хуавей - Україна»
Керівник напрямку Safe & Smart City,
к.т.н.

О.М. Дяденко
(підпис)
15.05.18
(дата)

КПІ ім. Ігоря Сікорського
Зав. кафедри ІТМ, д.т.н., проф.

Глоба Л.С.
Докторант кафедри ІТМ, к.т.н., с.н.с.

Скулиш М.А.
Асистент кафедри ІТМ

Суліма С.В.
15.05.18
(дата)